Regional identity of human neural stem cells determines oncogenic responses to histone H3.3 mutants

Regional identity of human neural stem cells determines oncogenic responses to histone H3.3 mutants

Raul Bardini Bressan[1,4], Benjamin Southgate[1,2], Kirsty M. Ferguson[1,2], Carla Blin[1], Vivien Grant[1], Neza Alfazema[1,2], Jimi C. Wills[2], Maria Angeles Marques-Torrejon[1], Gillian M. Morrison[1,2], James Ashmore[1], Faye Robertson[1,2], Charles A. C. Williams[1,2], Leanne Bradley[1,2], Alex von Kriegsheim[2], Richard A. Anderson[3], Simon R. Tomlinson[1,5] and Steven M. Pollard[1,2,*]

Authors Affiliations

[1] Centre for Regenerative Medicine and Institute for Regeneration and Repair, University of Edinburgh, Edinburgh EH16 4UU, United Kingdom.

[2] Cancer Research UK Edinburgh Centre, University of Edinburgh, Edinburgh EH4 2XR, United Kingdom.

[3] MRC Centre for Reproductive Health, Queens Medical Research Institute, University of Edinburgh, Edinburgh, EH16 4TJ, United Kingdom.

[4] Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen 2200, Denmark.

[5] Institute for Stem Cell Research, School of Biological Sciences, The University of Edinburgh, Edinburgh.

*** Corresponding and lead author:**

Steven Pollard: steven.pollard@ed.ac.uk Tel: +44(0)131 6519544

Running title: Regional identity determines responsiveness to H3.3 oncohistones

Keywords: neurodevelopment, cancer, cell-of-origin, histone H3.3, neural stem cells, paediatric high-grade glioma, glioblastoma, DIPG, ZMYND11

Summary

Point mutations within the histone H3.3 are frequent in aggressive childhood brain tumours known as paediatric high-grade gliomas (pHGGs). Intriguingly, different mutations arise in discrete anatomical regions: H3.3-G34R within the forebrain and H3.3-K27M preferentially within the hindbrain. The reasons for this contrasting aetiology are unknown. By engineering human foetal neural stem cell cultures from distinct regions, we demonstrate here that cell-intrinsic regional identity provides differential responsiveness to each mutant that mirrors the origins of pHGGs. Focusing on H3.3-G34R, we find that the oncohistone supports proliferation of forebrain cells, while inducing a cytostatic response in the hindbrain. Mechanistically, H3.3-G34R does not impose widespread transcriptional or epigenetic changes, but impairs recruitment of ZMYND11, a transcriptional repressor of highly expressed genes. We therefore propose that H3.3-G34R promotes tumorigenesis by stabilising the expression of key progenitor genes and, thus, locking initiating forebrain cells into their preexisting immature state.

Introduction

Genetic disruptions in chromatin-regulating pathways have been identified across many adult and paediatric tumours, thereby providing direct evidence for the critical role of epigenetic dysregulation in driving cancer (Filbin and Monje, 2019; Mohammad and Helin, 2017; Valencia and Kadoch, 2019; Weinberg et al., 2017). A seminal discovery in the field was the identification of recurrent mutations affecting histone H3 and epigenetic regulatory proteins in paediatric cancers (Schwartzentruber et al., 2012; Wu et al., 2012). Unique to pHGGs – universally lethal childhood brain tumours (Jones et al., 2016) – are mutations in *H3F3A,* which encodes the replication-independent histone H3.3 (Ahmad and Henikoff, 2002). Two single-copy mutations occur within the histone tail and typically give rise to substitution of either lysine 27 to methionine (K27M), or glycine 34 to arginine (G34R) (Schwartzentruber et al., 2012; Wu et al., 2012).

Intriguingly, K27M and G34R mutations are mutually exclusive and have a distinct anatomical distribution within the central nervous system (CNS); K27M is typically found in brainstem tumours (hindbrain-derived), while G34R occurs exclusively in hemispheric tumours (forebrain-derived) (Mackay et al., 2017; Sturm et al., 2012). Biochemically, H3.3-K27M has been shown to inhibit PRC2 (Polycomb Repressive Complex 2) (Lewis et al., 2013), while H3.3-G34R is thought to disrupt methylation of H3K36 (Lewis et al., 2013; Voon et al., 2018). However, the exact cellular and epigenetic mechanisms by which these 'oncohistones' operate during initiation of pHGGs and the reasons for their restricted anatomical specificity remain unknown (Filbin and Monje, 2019; Mohammad and Helin, 2017; Weinberg et al., 2017).

Here, we explore why these mutations arise with such anatomical precision. Of note, *H3F3A* is ubiquitously expressed across various tissues and we found no evidence of differential levels in the developing human CNS that could explain the biased distribution of the mutations (Figure S1). Based on the observation that H3.3 mutant tumours express region-specific neurodevelopmental signatures that correlate with their location (Mackay et al., 2017; Sturm et al., 2012), we hypothesised that transcriptional programs associated with regionalisation of the developing brain provide the critical cell-intrinsic vulnerabilities for each distinct H3.3 mutant.

To test this, we isolated primary human neural stem cells (NSCs) from forebrain and hindbrain regions of the developing human brain and engineered H3.3 and associated driver mutations. Using this model system, we demonstrate that hindbrain cells are more susceptible to H3.3-K27M oncogenic effects, whereas forebrain NSCs have a unique competence to respond to H3.3-G34R. We then focussed on H3.3-G34R – which has been less explored than H3.3-K27M – and explored its mechanism of action and reasons for the forebrain selectivity. Using both the engineered foetal NSC models and patient-derived cells, we demonstrate that H3.3-G34R does not induce pervasive transcriptional or epigenetic changes, but rather reinforces the preexisting progenitor state of the putative forebrain cell-of-origin. Mechanistically, we show that the G34R mutation impairs binding of H3.3 to the transcriptional repressor ZMYND11, which we propose acts focally to stabilise the expression of key forebrain progenitor genes, thereby locking cells into their immature state.

Results

Human foetal primary NSCs cultures retain their regional identity and recapitulate transcriptional signatures of pHGGs.

We reasoned that human foetal NSC cultures (Conti et al., 2005; Sun et al., 2008) derived from anatomically relevant regions could provide a tractable experimental system to model cell-intrinsic responses to H3.3 mutations. This allows investigation of early responses in an appropriate human cellular context, without the confounding influences of region-specific microenvironmental or broader systemic cues inherent to *in vivo* systems. We therefore derived primary NSC cultures from various regions of the developing human CNS (Figure 1A; see methods) and used these to assess the initial phenotypic responses to H3.3-K27M or H3.3-G34R oncohistones in vitro.

Established cultures uniformly expressed NESTIN, SOX2 and other pan-NSC markers (Figure S2A-B) and could be expanded using culture conditions that support self-renewal (Conti et al., 2005). Importantly, mRNA-profiling (RNA-seq) revealed that key forebrain and hindbrain transcriptional signatures that define the primary tissue samples were retained in the matched NSC cultures: forebrain-specific markers (e.g. *FOXG1*, *DLX2/6*, *EMX2*) were highly expressed in neocortex and striatum NSCs, while posterior markers (e.g. *IRX1/2* and *HOX* genes) were expressed in brainstem and spinal cord cultures (Figure 1B-D and Figure S2C).

To determine whether the region-specific foetal transcriptional signatures were relevant to pHGGs, we next compared our dataset with published primary tumour transcriptomes from three independent cohorts (Paugh et al., 2010; Schwartzentruber et al., 2012; Sturm et al., 2012). Indeed, through clustering approaches, the forebrain and hindbrain signatures successfully segregated the tumour samples based on H3.3 mutation status (Figure 1E-F; Table S1). Conversely, K27M- and G34R- tumour signatures distinguished the foetal tissue and matched NSC cultures based on their region-of-origin (Figure S2D-E; Table S1).

We next used single-sample gene set enrichment analysis (ssGSEA) to assess whether the pHGG subtype-specific signatures were specifically captured in the established NSC lines. Indeed, we observed enrichment of the G34R-associated signature in cultured neocortex and striatum NSCs,

while the defining K27M signature was preferentially enriched in brainstem and spinal cord cultures (Figure 1G-H). Consistent with this, neocortex and striatum cell line-specific gene sets were specifically enriched in G34R-mutant tumours, and brainstem and spinal cord NSC signatures in K27M tumours (Figure 1I). Thus, despite being expanded in identical culture conditions, NSC cultures largely retain their forebrain-hindbrain regional identities and recapitulate key transcriptional signatures that define pHGG subgroups.

Phenotypic responses to distinct H3.3 oncohistones mirrors their anatomical selectivity in pHGGs.

As the aforementioned results attested the relevance of our *in vitro* models for exploring pHGG aetiology, we next set out to evaluate the impact of cell-intrinsic regional identity on the cellular responses to H3.3 mutants. We first focussed on H3.3-K27M as a previous study using mouse models suggests preferential oncogenic activity of K27M in the hindbrain compartment (Larson et al., 2019). We induced stable expression of either wild-type (WT) or K27M mutant H3.3 using the PiggyBac transposon system in forebrain (neocortex or striatum) and hindbrain (brainstem) NSC cultures derived from two independent human foetal specimens (GCGR-NS19 and GCGR-NS13, see methods). A V5 epitope tag was fused to C-terminus of the construct to allow tracking of the ectopically expressed H3.3 protein (Figure 2A). Consistent with the reported biochemical activity of H3.3-K27M, we found substantial loss of global H3K27me3 in all lines, regardless of regional identity (Figure 2A). Phenotypically, however, overt oncogenic activity – i.e. increased proliferation and clonogenicity as well reduced senescence-associated βgalactosidase (SA-βgal) – was only observed in the brainstem, but not forebrain NSCs (Figure 2B-D). This indicates that global reduction of H3K27me3 alone is not sufficient to induce cellular responses and that H3.3-K27M tumour initiation may require a specific developmental cellular context. In a parallel study, we have characterised transcriptional and chromatin landscape changes in response to H3.3-K27M in these brainstem foetal NSCs (Brien and Bressan et al., submitted).

Given that region-specific cellular responses recapitulated the aetiology of H3.3-K27M pHGGs, we next focussed on understanding the less well studied H3.3-G34R mutant, for which mechanistic insights remain elusive and no models of tumour initiation are currently available to understand regional selectivity. Using the same experimental approach, however, we found that the G34R

mutant does not induce any overt proliferative effects in forebrain cultures (Figure 3A-E, Figure S3A-C). In contrast, in brainstem cultures, H3.3-G34R surprisingly triggered a cytostatic response with reductions in EdU incorporation and colony formation, as well as an increase in SA-βgal activity (Figure 3B-E). Together, our data indicate that, while forebrain cells can tolerate H3.3-G34R, hindbrain NSCs are intrinsically protected against its oncogenic effect, potentially explaining the absence of G34R-mutant tumours in posterior locations.

Due to the lack of overt phenotypic responses to H3.3-G34R in forebrain-derived NSCs, we reasoned that, even in an appropriate developmental context, cooperating mutations are required for manifestation of its oncogenic effects. We therefore introduced two additional oncogenic alterations commonly found in pHGGs: *PDGFRA* amplification and *TP53* loss-of-function (Mackay et al., 2017) (Figure 3F, hereafter referred to as PP5W or PP5G, for PDGFRA overexpression, P53 knockout and WT or G34R mutant H3.3, respectively). Indeed, compared to the WT H3.3 control (PP5W), neocortical PP5G cells displayed increased proliferation and clonogenicity, as well as a reduction in SA-βgal activity (Figure 3G). Upon transplantation into the cerebral hemisphere of immunocompromised mice, neocortical PP5G cells formed highly infiltrative neoplastic lesions confirming the anticipated role for H3.3-G34R in tumour initiation (Figure 3H-I). In stark contrast, engineering of PP5G mutations in brainstem NSCs triggered a strong cytostatic response compared to the PP5W control (Figure 3G) and did not render the cells tumour-initiating (data not shown). Similarly, combined engineering of *TP53* knockout and H3.3-G34R, without PDGFRA overexpression, also increased the proliferation of neocortical cultures, while inducing a cytostatic response in the brainstem counterpart (Figure S3D-F).

In summary, these findings demonstrate that, despite the identical culture conditions and similarities in broad NSC definition, regionally-distinct foetal NSCs display profound differences in their cell-intrinsic responses to oncohistones. Together, these highlight the importance of the regional identity in providing competence to distinct H3.3 mutations, which may explain the unique anatomical distribution of pHGGs in patients.

H3.3-G34R reinforces the transcriptional circuits that define forebrain identity

The preceding data suggested a role for H3.3-G34R in cooperation with other drivers within the forebrain. We next sought to investigate the effect of H3.3-G34R in fully transformed patient-derived tumour cells by eliminating the oncogenic allele and subsequent dissection of key downstream consequences in isogenic matched pairs. To that end, we used the CRISPR/Cas9 system to genetically inactivate the *H3F3A* G34R-mutant allele in a previously described mutant cell line (pGBM002) derived from a hemispheric (forebrain) patient tumour (Hennika et al., 2017) (Figure 4A, Figure S4A-B). Similarly to the primary forebrain NSC cultures, the patient cells uniformly expressed the NSC markers SOX2 and NESTIN and transcriptionally resembled the foetal forebrain. In addition, immunoblotting analysis indicated absence of ATRX protein, a H3.3 chaperone often mutated in G34R pHGGs (Mackay et al., 2017) (Figure S4A).

Following delivery of a mutant allele-specific sgRNA, five independent H3.3-G34R knock-out (G34R-KO) clonal lines were isolated and functionally characterised (Figure 4B-G). Compared to the parental cells, each displayed a similar phenotype, with slightly slower proliferation, reduced clonogenicity, and accumulation of SA-βgal-positive cells (Figure 4C-G). Importantly, *in vitro* overexpression of V5-tagged G34R-H3.3, but not WT-H3.3, was able to rescue the proliferative effect, thus validating the specificity of the knockout experiment (Figure S4C).

To dissect the mechanisms underlying the oncogenic effect of H3.3-G34R, we next defined the key transcriptional changes through mRNA-seq in the parental pGBM002 cells and G34R-KO clones (Figure 4I-L). Consistent with the cellular responses, principal component analysis confirmed separation of the G34R-KO clones as a distinct cluster from the parental cells (Figure 4I). Interestingly, we observed significant downregulation of many genes associated with forebrain development and neuroprogenitor proliferation in the G34R-KO cells, including *CDK6* (Mi et al., 2013), *SOX1/2* (Pevny and Placzek, 2005), *POU3F2/3* (Castro et al., 2006), *ARX* (Kitamura et al., 2002) and *DMRTA2* (Konno et al., 2012; Young et al., 2017) (Figure 4J-L). We identified only a marginal reduction in the expression of *MYCN*, a previously suggested target of H3.3-G34R based on differential expression analysis of non-isogenic tumour cell lines (Bjerke et al, 2013).

Analysis of transcriptomic data from patient tumour samples demonstrated that many of these forebrain-related genes were also more highly expressed in H3.3-G34R-mutant than H3.3-WT

tumours of hemispheric location (Figure 4M, Figure S4D). Altogether, these findings suggest that H3.3-G34R acts cell-autonomously to reinforce or stabilise a specific forebrain regulatory circuit present in the cell-of-origin. In agreement, we found that key transcription factor genes involved in neuroprogenitor self-renewal/proliferation were also upregulated in the engineered foetal neocortex model (PP5G vs PP5W cells, Figure S3G-H). These included the neuroprogenitor markers *OLIG2* and *SOX3*, as well as the forebrain-specific factors *EMX2*, *NR2F1*, *HOPX*, *DMRTA2* and *HIVEP2* (Bertacchi et al., 2020; Zweifel et al., 2018; Steinfeld et al., 2016; Ono et al., 2008; Yoshida et al., 1997). Of those, we note that *DMRTA2*, *EMX2, NR2F1* and *HIVEP2* were shared between engineered neocortex and patient-derived models, and may be part of a core regulatory network impacted by H3.3-G34R.

By contrast, no alterations in the expression of forebrain neural progenitor-related genes were observed in the engineered PP5W/G brainstem NSCs. Instead, transcriptional profiling indicated only modest global transcriptional changes (Figure S3H); however, most notably we observed a robust upregulation of the senescence effector and tumour-suppressor gene *CDKN1A* (encoding p21) (Abbas and Dutta, 2009). Intriguingly, no significant transcriptional changes were observed in the single mutant neocortex cells (G34R vs WT H3.3; Figure S3H), suggesting that manifestation of the H3.3-G34R effects at the transcriptional level might also require additional oncogenic signalling. Thus, we surmise that H3.3-G34R, in contrast to H3.3-K27M, requires additional oncogenic signalling for phenotypic activity and does not induce widespread transcriptomic changes. It instead appears to reinforce the pre-existing progenitor state of the forebrain cell-of-origin through increased expression of key transcriptional regulatory genes.

H3.3-G34R incorporates at pre-existing active genes and does not induce global changes in chromatin marks

We next capitalised on our patient-derived isogenic models to dissect the molecular mechanism underlying H3.3-G34R activity. In agreement with previously reported biochemical analyses (Lewis et al. 2013), we did not detect dramatic changes in total histone H3 post-translational modifications, supporting the idea that, unlike H3.3-K27M, H3.3-G34R does not act in a dominant manner to disrupt histone modification levels globally (Figure S5A).

To investigate whether G34R-H3.3 causes differential distribution of chromatin modification marks, we next mapped the genome-wide distribution of the mutant histone as well as active (H3K4me3/H3K36me3) and repressive (H3K27me3) histone marks through ChIP-Seq using parental and G34R-KO pGBM002 cells (Figure 5A). Similar to the reported distribution pattern of WT H3.3 at euchromatin (Goldberg et al, 2010), we observed higher H3.3-G34R enrichment at active promoters and gene bodies, which are marked respectively by high levels of H3K4me3 and H3K36me3 (Figure 5B-C, S5B-C). Consistently, H3.3-G34R ChIP reads directly correlated with mRNA expression levels, indicating greater levels of H3.3-G34R incorporation at highly transcribed genes (Figure 5D-E, S5D). Importantly, using the V5-H3.3 rescued G34R-KO cells, we found a strong correlation between WT and G34R V5-H3.3 ChIP signal (Figure 5F-G), indicating that the mutation does not grossly affect the H3.3 distribution patterns.

As altered patterns of genome-wide H3.3 distribution appear unlikely to explain the G34R transcriptional effects, we next sought to address whether incorporation of the mutant histone influenced the deposition patterns of the histone marks. Surprisingly, we did not find any significant correlation between G34R-H3.3 enrichment and changes in H3K4me3, H3K36me3 or H3K27me3 levels (Figure 5H, Figure S5C), indicating that the G34R mutant does not directly alter deposition of major histone marks as described for the other H3.3 oncohistones (Chan et al., 2013; Lewis et al., 2013; Fang et al., 2016; Lu et al., 2016; Khazaei et al, 2020). To further validate these observations, we next used ChIP peak calling to analyse the distribution of each histone mark at their cognate genomic regions; i.e. H3K4me3 and H3K27me3 at gene promoters and H3K36me3 at gene bodies (Figure 5I). In agreement, we found a large degree overlap of promoter-associated H3K4me3 and gene body-associated H3K36me peaks between parental and G34R-KO cells. Importantly, closer examination of the identified exclusive peaks (either parental or G34R-KO specific) indicated that those were low abundance peaks with only marginally lower levels in the reciprocal line (Figure 5I). Importantly, those differential bound peaks were not associated with consistent transcriptional expression changes of the underlying genes (Figure 5J). To a lesser extent, a high degree of overlap in repressive promoter-associated H3K27me3 peaks was also observed. Importantly, despite the detection of parental or G34R-KO specific peaks, these were

not associated with major transcriptional changes (Figure 5I-J). A similar pattern of genome-wide distribution of the investigated histone marks was also observed in the V5-H3.3 rescued G34R-KO cells (Figure S6).

These data indicate that, unlike H3.3-K27N and other H3.3 oncogenic forms, H3.3-G34R does not induce widespread epigenetic or transcriptional changes, and might instead operate through fine-tuning of preexisting expressed genes where it accumulates at highest levels.

G34R mutation disrupts binding of H3.3 to transcriptional repressor ZMYND11.

As H3.3 contributes to a small fraction of the total histone H3 pool within cells, and initial biochemical studies indicated that the G34R mutation prevents SETD2-mediated tri-methylation of the K36 residue on the mutant histone only (i.e *cis* loss of K36me3; Lewis et al., 2013; Zhang et al., 2017), we speculated that the mutation could impose transcriptional alterations through loss of H3.3K36me3 at specific loci. Using immunoprecipitation of the ectopically expressed V5-tagged H3.3 in G34R-KO rescued cells, we confirmed in a disease-relevant cellular context that the G34R mutation indeed causes loss of K36me3 on mutant histone but not on the total H3 pool (Figure 6A). In agreement, we found that the most downregulated genes in G34R-KO cells (i.e candidate G34R-upregulated targets) had higher than average H3.3-G34R coverage along their promoter and gene body regions (Figure 6B). Importantly, H3.3-G34R coverage appears particularly higher at the set of forebrain-related genes that are downregulated upon ablation of the G34R allele (Figure 6B-C), altogether indicating that H3.3-G34R acts mainly through transcriptional upregulation.

To address in an unbiased manner whether differential recruitment of biochemical partners may underlie the transcriptional changes imposed by H3.3-G34R, we next performed V5 epitope-based pull-down of WT- and G34R-H3.3 in the rescued G34R-KO cells and mass spectrometry to identify protein interaction partners (Figure 6D; Table S2). Strikingly, we found that H3.3-G34R has reduced binding to ZMYND11, a specific reader of H3.3K36me3 that functions as an unconventional transcription co-repressor by modulating elongation and splicing of highly

expressed genes (Wen et al 2014; Guo et al 2014). Reduced binding to ZMYND11 was also confirmed by Co-IP experiments in both patient-derived and engineered forebrain NSC isogenic cell line pairs (Figure 6E).

Further supporting the involvement of ZMYND11, we found using computational analyses that genes preferentially affected by H3.3-G34R (including the set of key forebrain-associated genes) collectively display sequence features, such as higher CpG content and presence of intragenic CpG islands (Figure 6F), that are associated with the control of transcriptional elongation and splicing (Veloso et al., 2014, Monteuuis et al., 2019). Altogether, these data support a model in which G34R-induced loss of H3.3K36me3 at highly expressed loci (including forebrain master regulator genes such as *DMRTA2* and *ARX*) precludes recruitment of the transcriptional repressor ZMYND11. We surmise this ultimately results in elevated transcriptional output of such genes and helps to sustain the core regulatory network of forebrain identity. Due to the presence of high H3.3 incorporation and unique gene sequence features (e.g high intragenic CpG content), we speculate that the altered genes may have specific requirements for elongation and/or splicing control mechanisms that explain their increased sensitivity to disruption of ZMYND11 recruitment.

FOXG1 is a competence factor that enables H3.3-G34R driven transformation.

Surprisingly, despite the transcriptional changes and *in vitro* phenotype conferred by H3.3-G34R, we found that G34R-KO patient cells were still able to form tumours upon xenotransplantation (Figure 4H), suggesting that the mutation might be dispensable later in tumorigenesis, once widespread genomic disruptions have accumulated. Given the cytostatic effect triggered by H3.3-G34R in non-forebrain cells, we reasoned that targeting components of the forebrain transcriptional network could provide an alternative strategy to limit tumorigenicity of fully transformed cells. Interestingly, we observed that the cytostatic effect induced by H3.3-G34R in brainstem NSCs is accompanied by activation of the senescence effector *CDKN1A*/p21 (Figure S7A-B), suggesting that cellular senescence is a key protective mechanism that can be overcome in forebrain cells.

We therefore focussed on the transcription factor FOXG1, a master regulator of forebrain identity and known suppressor of the *CDKN1A*/p21 locus during brain development (Xuan et al., 1995; Seoane et al., 2004; Siegenthaler et al., 2008; Yip et al., 2012; Lujan et al. 2012). Two other aspects of FOXG1 biology were relevant. First, *FOXG1* is the most upregulated transcription factor gene in forebrain versus hindbrain NSC cultures (Figure 1B). Second, *FOXG1* is expressed at high levels in hemispheric pHGGs independently of the H3.3-G34R mutation (Figure 7A, Figure S7C-D) and in many adult HGGs (Verginelli et al., 2014; Bulstrode et al. 2017).

Strikingly, ablation of *FOXG1* through CRISPR/Cas9-gene editing led to reduced proliferation with concomitant p21 upregulation *in vitro* (Figure 7B-D) and loss of tumour-formation capacity of H3.3-G34R patient-derived cells *in vivo* (Figure 7E-G). In contrast, ablation of *FOXG1* in two independent H3.3-WT adult hemispheric HGG patient lines with comparable expression levels did not induce p21 upregulation nor impaired proliferation, suggesting a stronger dependency of H3.3-G34R-mutant cells on *FOXG1* expression (Figure 7H-L). We therefore propose that FOXG1 acts as a regionally-restricted competence factor for H3.3-G34R via establishment of the permissive forebrain transcriptional state and/or direct protection against H3.3-G34R-induced p21 activation.

Discussion

We have demonstrated here the impact of regional NSC identity on the oncogenic effects of H3.3 oncohistones, and propose that this can explain the unique anatomical distribution of high-grade brain tumours in children. Our findings suggest that cell-intrinsic programs associated with regional identity of human neural progenitor cells provide unique vulnerabilities to each H3.3 mutant form. Thus, rather than region-specific microenvironmental influences or systemic temporal cues, our data point to the primary importance of the regional identity of the cell-of-origin in providing a cell intrinsic competence to distinct H3.3 mutations.

The differing cellular responses to H3.3-K27M and H3.3-G34R H3.3 oncohistones clearly highlight the importance of studying these mutations in an appropriate spatial-temporal developmental context. Understanding of basic disease biology therefore requires modelling in human cells with the relevant regional identity with associated transcriptional circuits and epigenetic landscapes. Our isogenic series of forebrain and hindbrain self-renewing NSCs now provides a human-relevant model to further dissect cellular responses to oncohistone and the molecular events underlying initiation of pHGGs. These should complement patient-derived models, which are often too corrupted and heterogeneous to retrospectively delineate programs of early oncogenesis. At a practical level, these isogenic models can be expanded continuously in large numbers, are amenable to genetic manipulation (Bressan et al., 2017) and, similar to adult glioma stem cell lines (Pollard et al., 2009), represent a tractable experimental system for chemical and genetic screens in search of new therapeutic targets in pHGGs.

Importantly, working with tissue-derived lines overcomes many of the shortcomings associated with ES/iPSC-based models, which are often highly capricious, require lengthy differentiation protocols, and result in heterogeneous mixtures of cell types. Noteworthy, in addition to dissecting the role of H3.3 oncohistones, the defined origin and identity of our primary foetal NSC lines may also facilitate modelling of other paediatric tumors that arise in specific CNS compartments, such as medulloblastoma or ependymoma (Gibson et al., 2010: Johnson et al., 2010). In light of this, we and others are developing a large collection of human foetal NSC lines across different

temporal and spatial identities that should enable functional dissection of diverse brain tumour drivers, and are being made available to the research community (www.gcgr.org.uk).

Further leveraging on our isogenic models, we were able to garner molecular insights into the H3.3-G34R mode of action in the relevant cellular context. Through a combination of mRNA-seq profiling and mapping of histone post-translational modifications, we observed that H3.3-G34R does not induce pervasive transcriptional or epigenetic changes, but rather operates at already highly expressed genes, including key forebrain NSC transcription factors. These findings clearly contrast to the reported effects of H3.3-K27M in midline pHGGs and other H3.3 driver mutations present in paediatric sarcomas, where the oncohistones appear to cause widespread resetting of epigenetic landscapes (Chan et al., 2013; Fang et al., 2016; Lu et al., 2016; Khazaei et al, 2020)

In addition, by exploiting the epitope tagged version of H3.3, we were able to map key protein interaction partners of H3.3-G34R in patient-derived cells. Consistent with the idea of alterations at specific loci, we identified disrupted binding to ZMYND11 - a putative tumour suppressor that is associated with highly expressed genes and operates as a transcriptional co-repressor through modulation of elongation and splicing. Interestingly, a previous study has also identified a clonal ZMYND11 loss-of-function mutation in hemispheric pHGGs lacking H3.3-G34R, suggesting they may operate through a shared pathway (Salloum et al., 2017).

We therefore propose a working model in which H3.3-G34R does not induce *de novo* changes in transcriptional state, but instead acts to fine tune expression of key transcriptional regulatory genes via the release of ZMYND11-based transcriptional repression. Ultimately, this reinforces the pre-existing progenitor state of the forebrain cell-of-origin, and likely locks cells into a proliferative state. Although mechanistically different, these findings add to emerging evidence that mutations in histone and chromatin regulatory genes captures susceptible stem cell and progenitor states to drive initiation of paediatric tumours (Funato et al., 2014; Lu et al., 2016; Jessa et al., 2019; Pathania et al., 2017; Vladoiu et al., 2019, Khazaei et al., 2010). Further studies will now be needed to deeply assess the impact of H3.3-G34R on transcriptional elongation and/or alternative splicing, and whether its effect can be fully recapitulated by ZMYND11 loss-of-function. Similarly, in-

depth biochemistry and structural analyses will be required to dissect whether ZMYND11 binding is directly impaired by the G34R substitution, or rather an indirect consequence of K36me3 loss on the mutant histone.

Finally, our data suggests that targeting core lineage-affiliated transcription factors, such as FOXG1, might be an attractive approach for therapeutic development. While the design of conventional small molecule inhibitors that target transcription factors presents considerable challenges, there has been significant progress in targeting these with innovative strategies, including modulation of auto-inhibition, proteolysis targeting chimaeras (PROTACs), and cysteine reactive inhibitors (Bushweller, 2019). Increased understanding of transcription factor biochemical functions may further reveal critical post-translational modifications and/or co-factors that could be disrupted for therapeutic purposes. We also note the shared dependencies on FOXG1 for both paediatric H3.3-G34R-mutant and adult HGGs (Bulstrode et al., 2017; Verginelli et al., 2013). It appears that, despite being driven by divergent genetic disruptions (Jones et al., 2017), both adult and paediatric diseases may converge and harness elevated expression of a forebrain neural stem/progenitor cell transcriptional network. It will be interesting in future studies to explore the role of other key components of the forebrain regulatory circuit and to dissect in detail how these are established and impacted by H3.3-G34R.

In conclusion, our data indicate that the regional identity of human NSCs is critically important for the oncogenic responses to distinct H3.3 mutant forms. Further work will now be needed to fully dissect why regional progenitor states are differentially sensitised to each H3.3 mutation, and whether mechanisms sustaining forebrain and hindbrain identities could be targeted for therapeutic purposes. Our hope is that this identification of the crucial target cell populations, and underlying mechanisms of susceptibility, may ultimately lead to more effective therapies for this devastating disease.

Limitations of the study

A current major limitation in studies of pHGG is the difficulty in obtaining tumour specimens and patient cell lines due to the challenging anatomical location and rarity of the disease. Although our findings with the patient-derived pGBM002 cell line are supported by the engineered foetal NSC models and rescue experiments, it will be important to further validate key target genes and the requirement of the mutant histone in late tumorigenesis as novel models emerge and other patient lines became available to the community. Similarly, while we clearly demonstrate the importance of the forebrain developmental context, it will be important to determine in future studies whether dorsal-ventral regional programs or progenitor subtype identities provide further layers of competence to H3.3-G34R. Mechanistically, deeper analyses of the total RNA expression will be required to explore the impact of H3.3-G34R on elongation and splicing control, and whether ZMYND11 inhibition recapitulates the effect of H3.3-G34R. Finally, in exploring the role of FOXG1, we have been unable to isolate expandable cells following conventional knockout and overexpression approaches in forebrain and hindbrain NSC cultures, respectively. Modulation of *FOXG1* expression using CRISPRa/CRISPRi might provide an alternative experimental strategy to further delineate its role as a competence factor.

Acknowledgements

We thank Adrian P. Bracken, Katrin Ottersbach, Kim B. Jensen and members of the Pollard lab for helpful discussions and critical reading of the manuscript. Abdenour Soufi provided technical advice on ChIP-seq protocols. We are grateful to the University College London Cancer Institute Genomics core facility for performing the next generation mRNA sequencing, and to the Wellcome Trust for providing the mass spectrometry large equipment (Multiuser Equipment Grant 208402/Z/17/Z). Primary foetal NSCs were provided by the Glioma Cellular Genetics Resource (www.gcgr.org.uk). Angel Montero Carcaboso (SJD Barcelona Children's Hospital, Spain) generously provided the patient-derived cell line pGBM002 (HSJD-GBM-002). The project was supported by a grant from the Children with Cancer UK Charity (15/189). R.B.B. was supported by a PhD fellowship from the Science Without Borders Program (CAPES, Governo Dilma Rousseff, Brazil) and a postdoctoral fellowship from EMBO. G.M.M. was supported by the Cancer Research UK Centre Accelerator Award (A21922). S.M.P is a Cancer Research UK Senior Research Fellow (A17368).

Author Contributions

R.B.B and S.M.P conceived the project and designed the experiments. R.B.B. performed the majority of the bench-based experimentation and data analyses. B.S., J.A. and S.T. performed the bioinformatic analyses. N.A., L.B. and M.A.M-T. performed xenotransplantation experiments and histological analyses. C.B. performed the immunoprecipitations and western blotting procedures. J.W. and A.K. performed mass-spectrometry analysis. K.M.F and F.R. performed phenotypic characterization of FOXG1 loss-of-function experiments. V.G, C.W., K.M.F. and G.M.M. provided technical assistance with derivation and characterization of the cell lines, and R.A.A provided human foetal tissues. R.B.B. and S.M.P. co-wrote the manuscript.

Declaration of interests

S.M.P is a founder and shareholder of Cellinta Ltd., a biotechnology startup that is developing cancer therapeutics, including glioblastoma. He is also a paid consultant to Cellinta Ltd. S.M.P is

an inventor on a University of Edinburgh patent related to neural stem cell culture methods (WO2005121318A3). The other authors declare no competing interests.

Figure legends

Figure 1 | Core transcriptional signatures of regional identity are captured in human primary foetal NSC cultures and define pHGG subtypes. See also Figure S2 and Table S1

(A) Overview of the approach used for derivation of human foetal region-specific NSC cultures. Three foetal specimens aged 13.5, 15.0 and 19.3 WGA (GCGR-NS13, GCGR-NS15 and GCGR-NS19, respectively) were used. Established lines were profiled after 5-10 passages in culture. Total RNA collected from the freshly dissected tissue segments was used as a reference control. WGA, weeks of gestational age.

(B) Heatmap showing expression of forebrain- and hindbrain-specific genes in the freshly dissected primary tissue and matched NSC cultures. Region-specific expression signatures were generated for reference tissue samples and NSC cultures separately, and the intersection of the two sets was taken (see Methods). Forebrain- and hindbrain-specific gene symbols are shown in blue and green lettering, respectively.

(C) Principal component analysis (PCA) of primary foetal tissue samples and NSC lines based on 500 genes with the highest variance across all samples.

(D) Gene ontology analysis (domain biological process) of PCA loadings for principal component 1 (PC1) and principal component 2 (PC2). PC1, which segregates NSC lines from primary tissue samples, is enriched in neuronal differentiation terms. In contrast, PC2 separates samples according to anatomical region of origin and is enriched in regionalisation and antero-posterior specification terms.

(E) PCA of primary patient pHGG tumour samples (GEO accessions: GSE34824, GSE36245, GSE19578). Left panel: PCA based on tumour-derived G34- and K27-mutant signatures. Right panel: PCA based on forebrain- and hindbrain-specific gene sets extracted from the reference foetal tissue and NSC cultures in panel (A).

(F) Consensus clustering of patient pHGG tumour samples using G34/K27 (left) and forebrain/hindbrain (right) gene sets. Heatmaps are clustered using hierarchical clustering with Euclidean distance and average linkage criteria into 3 predominant groups corresponding to tumour H3.3 status (WT, K27M, and G34R/V).

(G) Boxplots of ssGSEA enrichment scores for K27 and G34 signatures among the established region-specific foetal NSC lines. Cell line numbering corresponds to the anatomical regions in

panel (a). Boxes show median, first and third quartile, and whiskers extend to 1.5× the interquartile range (IQR).

(H) Boxplots of ssGSEA enrichment scores comparing forebrain and hindbrain signatures extracted from the primary foetal tissue datasets among pHGG H3.3 subtypes. Boxes show median, first and third quartile, and whiskers extend to 1.5× the IQR.

(I) Boxplots of ssGSEA enrichment scores comparing the gene sets extracted from the established NSC lines from each position among pHGG H3.3 subtypes. Boxes show median, first and third quartile, and whiskers extend to 1.5× the IQR.

Figure 2 | H3.3-K27M preferentially increases growth of brainstem NSCs.

(A) Immunocytochemistry analysis confirming comparable levels of V5-tagged WT and K27M H3.3 expression in the three regional NSC lines. Global loss of H3K27me3 is observed in H3.3-K27M-expressing cells. Representative images of the lines derived from specimen GCGR-NS13 are shown. Scale bar: 20 μm.

(B) Quantification of EdU incorporation in the regional NSC lines carrying H3.3-WT and -K27M constructs. Plots depict mean ± standard deviation percentage EdU-positive cells in the indicated cell lines. Student's t-test p-values are shown. n.s, non-significant. n = 3 independent experiments performed with 3-4 technical replicas each.

(C) Quantification of colony formation activity of regional NSC lines carrying H3.3-WT and -K27M constructs. Individual dots represent the absolute colony counting per well. Horizontal lines represent mean ± standard deviation. Student's t-test p-values are shown.

(D) Quantification of SA-βgal positive cells in the regional NSC lines carrying H3.3-WT and -K27M constructs. Plots depict mean ± standard deviation percentage of positive cells**.** Student's t-test p-values are shown. n = 3 independent experiments performed with 3-4 technical replicas each.

Figure 3 | H3.3-G34R triggers differential phenotypic responses in forebrain and brainstem NSCs. See also Figure S3

(A) Immunoblot analysis confirming comparable expression of V5-tagged H3.3 in the NSC lines carrying WT and G34R H3.3 constructs. Representative blot images from GCGR-NS13 lines are shown. Expression of forebrain marker FOXG1 is shown for comparison purposes.

(B) Quantification of EdU incorporation in the regional NSC lines expressing H3.3-WT and -G34R. Plots depict mean ± standard deviation percentage of positive cells in the indicated cell lines. Student's t-test p-values are shown. n.s, non-significant. n = 3 independent experiments performed with 3-4 technical replicas each.

(C) Quantification of colony formation activity of regional NSC lines expressing H3.3-WT and -G34R. Individual dots represent the absolute colony counting per well. Horizontal lines represent mean ± standard deviation. Student's t-test p-values are shown.

(D) Quantification of apoptotic, active Caspase3-positive cells in the regional NSC lines carrying H3.3-WT and -G34R constructs. Plots depict mean ± standard deviation percentage of positive cells**.** p-value by Student's t-test. n = 3 independent experiments performed with 3-4 technical replicas each.

(E) Quantification of SA-βgal positive cells in the regional NSC lines expressing H3.3-WT and -G34R. Plots depict mean ± standard deviation percentage of positive cells**.** Student's t-test p-values are shown. n.s, non-significant. Representative images of indicated lines derived from specimen GCGR-NS13 are shown. Scale bar: 100 μm. n = 3 independent experiments performed with 3-4 technical replicas each.

(F) Immunoblot analysis confirming loss of P53 protein, and ectopic expression of V5-tagged H3.3 and PDGFRa in the engineered GCGR-NS13 Neocortex and Brainstem PP5W and PP5G cells. Protein lysates from non-edited cells were used as controls.

(G) Phenotypic *in vitro* characterisation of GCGR-NS13 neocortex and brainstem PP5W and PP5G cells. Plots show quantification of EdU incorporation (n =3), colony formation (n = 12) and SA-βgal assays (n =3).

(H) Kaplan-Meier curve showing survival data of immunocompromised mice transplanted with GCGR-NS13 neocortical PP5W and PP5G cells.

(I) Analysis of brain tissue from mice transplanted with GFP-labelled neocortical PP5W and PP5G cells. Left panel shows microscopic views of freshly dissected brains from each group. Right panel shows immunohistochemistry images (coronal sections) of the corresponding specimens. The transplanted human cells are stained with a GFP antibody. Scale bar: 1.5mm

Figure 4 | Genetic ablation of mutant *H3F3A* allele in patient-derived NSC line reveals forebrain neuroprogenitor genes as downstream targets of H3.3-G34R. See also Figure S4

(A) Schematics of the strategy used to inactivate the *H3F3A* G34R allele through CRISPR/Cas9 gene editing. The G-to-A nucleotide substitution that results in the G34R mutation occurs 5' adjacent to a GGG stretch (red) that serendipitously serves as a protospacer adjacent motif (PAM) for Cas9 recognition. As mismatches within the PAM-proximal region preclude Cas9 cleavage, this allows the design of a sgRNA (blue sequence) that specifically cuts the G34R allele. Dashed rectangles show codons for residues Gly34 (WT allele) or Arg34 (G34R allele).
(B) Genotyping result of the parental and a representative G34R-KO clone. WT allele (encoding Gly34) remains intact, while G34R allele harbours a frameshifting 1 bp deletion at the predicted sgRNA cut site (yellow arrow).
(C) Immunocytochemistry analysis confirms loss of H3.3-G34R protein expression in the G34R-KO cells. Parental line and representative G34R-KO clone are shown. Scale bar: 50 μm.
(D) Confluence analysis indicates impaired *in vitro* growth of G34R-KO clones. Parental cells were used as reference. **** adjusted $p < 0.0001$ by one-way ANOVA followed by Dunnett's test (each G34R-KO clone vs parental cells at 240 h elapsed). Plot shows results of one representative experiment performed with 4 technical replicas.
(E) Quantification of EdU incorporation in pGBM002 parental and G34R-KO clonal lines. Plot depicts mean ± standard deviation. *$p < 0.05$, ***$p<0.001$ by One-way ANOVA followed by Dunnett's multiple comparisons test (mean of each clone vs parental). $n = 4$ independent experiments performed with 3-4 technical replicas each.
(F) Quantification of colony formation activity of parental pGBM002 and G34R-KO clonal lines. Individual dots represent the absolute colony counting per well. Horizontal lines represent mean ± standard deviation. ****$p<0.0001$ by One-way ANOVA followed by Dunnett's multiple comparisons test (mean of each clone vs parental).
(G) Quantification of SA-βgal positive cells in parental pGBM002 and G34R-KO clonal lines. Plot depicts mean ± standard deviation. *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$ by One-way ANOVA followed by Dunnett's multiple comparisons test (mean of each clone vs parental). $n = 3$ independent experiments performed with 3-4 technical replicas each.
(H) Results of xenotransplantation experiments with parental pGBM002 and G34R-KO clones. Left panel shows representative images of *in vivo* bioluminescence imaging of intracranial tumours formed by parental and a G34R-KO clonal line. Kaplan-Meier curve (right) shows survival data of mice transplanted with the parental cells (blue line) and each G34R-KO clone (grey lines).

(I) PCA plot confirms consistent transcriptional profile changes between parental pGBM002 (three independent passages) and three G34R-KO clonal lines (one passage each). The 500 genes with highest variance across samples were used for the PCA.

(J) Volcano plot showing differentially expressed genes between parental and G34R-KO clonal lines. Genes downregulated in the G34R-KO clones are shown in blue. Upregulated genes shown in red. Vertical dotted lines mark absolute $\log_2$ fold change > 1. Horizontal lines - false discovery rate (FDR) < 0.05.

(K) Gene ontology (GO) analysis of downregulated and upregulated genes (FDR < 0.05) in G34R-KO clones for GO domain Biological Process.

(L) Heatmap showing expression of genes within the 'Forebrain Development' GO term found to be downregulated in G34R-KO clones.

(M) Boxplot showing expression levels of the indicated forebrain progenitor-related genes in hemispheric pHGG patient tumour samples. Data was retrieved from an integrated transcriptomic study of pHGGs available on the paediatric cBioPortal – see methods). Only specimens whose hemispheric location and *H3F3A* mutation status could be ascertained were included. * p < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001 by Student's t test (WT vs G34R/V groups).

Figure 5 | Genome-wide profiling of H3.3 and associated histone H3 marks in isogenic patient-derived H3.3-G34R pHGG cells. See also Figures S5-S6.

(A) Heatmaps and metagene plots showing H3.3-G34R, H3K4me3, H3K36me3 and H3K27me3 ChIP-seq signal at all annotated human genes (hg38 knownGene) in parental and G34R-KO pGBM002 cells. Genes within each heatmap are ranked by abundance of the H3.3-G34R or indicated histone mark in the parental cells.

(B) Box plots depicting abundance of each indicated histone H3 modification ($\log_2$ RPKM) at genes grouped into quantiles of G34R abundance in parental pGBM002 cells. Quantiles 1 and 4 include, respectively, genes with the lowest and highest H3.3-G34R coverage levels. Boxes show median, first and third quartile, and whiskers extend to 1.5× the IQR (n=2 replicates for each histone mark).

(C) Box plots depicting abundance of H3.3-G34R ($\log_2$ RPKM) at genes grouped into coverage quantiles of the indicated histone mark. Quantiles 1 and 4 include genes with the lowest and highest levels of each mark, respectively (n=2).

(D) Box plots depicting total mRNA expression levels (log_2 RPKM) per H3.3-G34R quantile in parental and G34R-KO cells. Quantiles 1 and 4 include, respectively, genes with the lowest and highest H3.3-G34R coverage levels (n=3).
(E) Box plot depicting H3.3-G34R coverage (log_2 RPKM) in parental cells per mRNA expression quantiles defined in parental (left) and G34R-KO (right) cells. Quantiles 1 and 4 include, respectively, genes with the lowest and highest absolute mRNA expression levels (n=2).
(F) Scatter plot showing V5-H3.3 coverage per 50kb genomic bins in G34R-KO cells rescued with V5-tagged H3.3 WT - and G34R construct. Density reported as log_2 density. SCC 0.91, p-value < 0.001.
(G) Tracks of V5-WT and V5-G34R H3.3 subtracted and normalised coverage (RPM) over chromosome 8 (upper), and selected region (lower, Chr8 p22-21.2).
(H) Log_2 fold change (Parental vs G34R-KO cells) in H3K4me3, H3K36me3, and H3K27me3 coverage (RPKM) within each H3.3-G34R quartile.
(I) Venn diagrams and metagene plots of common and exclusive peaks in pGBM002 Parental and G34R-KO lines (Left to right – Parental, Common, and G34R-KO peaks). Peaks were selected as promoter specific (+/- 2kb from TSS) for H3K4me3 and H3K27me3 histone marks, and gene body associated (between TSS and TES) for H3K36me3.
(J) RNA-seq mRNA log_2 fold change (Parental Vs. G34R-KO) of genes associated with peaks defined in panel (I).

Figure 6 | G34R mutation precludes binding to the transcriptional repressor ZMYND11. See also Table S2
(A) Immunoblotting of anti-V5-immunoprecipitated H3.3-WT and -G34R ectopically expressed in patient derived G34R-KO pGBM002 cells. Endogenous (lower band) and V5-tagged ectopic H3.3 (upper band) are indicated. Loss of K36me3 is specific to ectopic H3.3-G34R. Non-transfected G34R-KO cells were used as controls (first and fourth lane).
(B) Metagene plot (left) and boxplot (right) showing H3.3-G34R coverage (subtracted RPM, and log_2 RPM respectively) for Expressed (n=13695), Upregulated (n=842), Downregulated - all (n=475), and Downregulated – forebrain associated (n=44) genes. Expressed genes were defined as those with mean RPKM > 1 within pGBM002 parental samples (see methods). All other gene sets were defined based on differential expression between G34R-KO and parental cells (Fig.4I-

L). *** $p < 0.005$ and ****$p < 0.0001$ for pair-wise comparisons against the 'all expressed' group. See STAR methods for details on the statistical test.

(C) Genomic track showing H3.3-G34R enrichment and mRNA expression over transcriptional factor genes *DMRTA2* (left) and *DLX1* (right) in parental and G34R-KO pGBM002 cells. *DMRTA2* and *DLX2* are shown as representative forebrain-related downregulated and unchanged genes in G34R-KO cells.

(D) Volcano plot showing differentially bound partners of ectopically expressed WT and G34R mutant V5-tagged H3.3 identified through IP-MS. The x-axis shows the fold-change value (G34R/WT) and y axis shows p-value (−log10 scale) for statistical significance. Absolute 3-fold change and p-value 0.1 (vertical and horizontal dashed lines, respectively) were used as the threshold cutoff. Blue and green dots depict partners with increased and decreased binding to H3.3-G34R, respectively.

(E) Immunoblotting analysis for ZMYND11 and HIRA co-immunoprecipated with ectopic V5-H3.3 (WT and G34R) in G34R-KO pGBM002 patient cells, and engineered neocortex NSC line (H3.3 overexpression alone, and in combination with p53-knockout and PDGFRA overexpression - PP5W/PP5G model).

(F) Boxplots showing quantification of sequence length, GC content, CpG content, and CpG island overlap for full transcript, exonic and intronic ranges in the defined gene sets. Only expressed genes were included in the analyses - see methods for details. All expressed (n=13695), Downregulated - all (n=475) and Downregulated – forebrain associated (n=44). ** $p < 0.05$, *** $p < 0.005$ and ****$p < 0.0001$ for pair-wise comparisons against the 'all expressed' group. See STAR methods for details on the statistical test.

Figure 7 | Genetic disruption of FOXG1 impairs tumourigenesis of patient-derived G34R-mutant NSC cells. See also Figure S7

(A) Expression levels of *FOXG1* mRNA in patient pHGG primary tumours. Data was retrieved from the integrated pHGG transcriptomic dataset available on the paediatric cBioPortal – see methods. N.s, non-significant and **** $p < 0.0001$ by one-way ANOVA followed Tukey's multiple comparisons test.

(B) Immunoblot analysis confirming loss of FOXG1 protein in the pGBM002 knockout clonal line. Blot against GAPDH was used as internal control.

(C) Quantification of EdU incorporation (24h pulse) and colony formation activity in parental and FOXG1-knockout pGBM002 cells. Plot depicts mean ± standard deviation. **** $p<0.0001$ by Student's t test. n = 4 independent experiments performed with 3-4 technical replicas each.

(D) Immunoblot analysis showing increased p21 expression in the FOXG1-knockout clonal line following exposure to TGFβ in the culture media for 96 h. TGFβ treatment was used to stabilise p21 protein expression as previously described (Seoane et al., 2014).

(E) *In vivo* bioluminescence imaging of mice transplanted with parental and FOXG1-knockout pGBM002 cells labelled with a constitutively expressed GFP-Luciferase reporter. Plot depicts absolute luminescence values (photons per second) per mouse at the indicated time points. Representative images were collected at 150 days following transplantation.

(F) Representative immunohistochemistry images of the brain tissue from immunocompromised mice transplanted with parental and FOXG1-knockout pGBM002 cells.

(G) Kaplan-Meier curve showing survival data of mice transplanted with parental and FOXG1-KO pGBM002 cells. Mice in the FOXG1-KO group were sacrificed after 400 days and no signs of tumour formation was observed. Scale bar: 1.5mm

(H) RT-qPCR-based relative quantification of *FOXG1* expression levels in H3.3-G34R mutant pHGG (pGBM002) and H3.3-WT adult HGG (aGBM7 and aGBM313). Primary foetal neocortex and brainstem NSC lines (specimen GCGR-NS19) were used for comparison. n = 2-3 independent passages of each cell line.

(I) Western blotting analysis showing protein expression of FOXG1 and p21 in parental and respective FOXG-KO lines. H3.3 status of each line is shown in the round rectangles below the panel.

(J) RT-qPCR results showing relative expression levels of *CDKN1A* in FOXG1-KO cells. Parental cells were used as normaliser for each pairwise comparison. **$p < 0.01$ by Student's t-test. n.s, non significant. n = 3 independent passages of each cell line.

(K) Quantification of EdU incorporation (2h pulse) in FOXG1-KO and respective parental lines. Plots depict mean ± standard deviation. *** $p<0.001$ by Student's t-test. n.s, non significant. n = 3-4 independent experiments performed with 3 technical replicas each.

(L) Quantification of SA-β-gal positive cells in FOXG1-KO and respective parental lines. Plots depict mean ± standard deviation. *$p < 0.05$ by Student's t-test. n.s, non significant. n = 4 independent experiments performed with 3 technical replicas each.

STAR Methods

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Steven Pollard (steven.pollard@ed.ac.uk).

Materials availability

Stable reagents generated in this study are available from the Lead Contact. Unedited human foetal human NSC lines must be requested from the Glioma Cellular Genetics Resource (gcgr.org.uk) and a Materials Transfer Agreement completed.

Data and Code Availability

The RNA-seq and ChIP-seq datasets reported in this paper are deposited in Gene Expression Omnibus (GEO) with the following accession codes: GSE163044 and GSE163185. Proteomics data is available at ProteomeXchange with identifier PXD021270.

Plasmids

H3.3 expression vectors were generated through engineering a Gateway destination vector containing a CAG promoter-driven transcriptional unit and a PGK-Hygromycin selection cassette flanked by PiggyBac transposon-specific inverted repeats. The sequences encoding C-terminal V5-tagged WT, K27M- and G34R-mutant H3.3 were generated by commercial DNA synthesis (Life Technologies). These were flanked by AttL recombination sites and directly inserted into the destination vector via Gateway LR cloning. PDGFRA coding sequence was PCR-amplified from human cDNA and cloned into a similar Gateway-compatible vector containing a CAG-Blasticidin resistance cassette and PiggyBac transposon flanking sequences. The PiggyBac Luciferase-2A-GFP expression vector was generated through LR Gateway cloning of an existing entry clone (Bressan et al. 2017). Human *TP53* and *FOXG1* CRISPR sgRNAs and targeting vectors were previously described (Bressan et al. 2017; Bulstrode et al. 2017). Cas9-encoding plasmid was obtained from Addgene. PBase transposase plasmid was a gift from Austin Smith (University of Cambridge).

Human tissue collection and derivation of NSC cultures

Fresh human foetal CNS tissue was obtained following elective termination of pregnancy at 13.5, 15.0 and 19.3 weeks of gestational age (WGA) (specimens GCGR-NS13, GCGR-NS15 and GCGR-NS19, respectively) with no evidence of developmental abnormalities. Informed consent was obtained in writing, and all procedures received ethical approval (South East Scotland Research Ethics Committee, REC reference 08/S1101/1). The regions of interest (Figure 1a) were dissected under a stereoscope with a pair of fine forceps in ice-cold PBS. In the case of the brainstem, an incision was made at the midbrain-hindbrain boundary and medullary hindbrain-spinal cord boundary. Following dissection of the cerebellum, an incision was made at the level of the rhombic lip, thereby dividing the segment into anterior and posterior brainstem. In the case of the spinal cord, the proximal segment (approx. 3-5 mm adjacent to the medulla) was collected. Following dissection, each tissue segment was split in two, with the first half (tissue reference) being directly lysed in Qiagen RLT plus buffer for subsequent RNA collection. The second half was mechanically dissociated with a pair of forceps followed by gentle pipetting. The resultant cell suspensions were centrifuged (400g for 3min), resuspended in NSC culture medium (see below) and plated onto T25 tissue culture flasks. Within 24h, morphologically diverse cells attached to the culture plastic. To reduce the amount of debris and cell death, culture media was refreshed at 24h and then every seven days. Within 1-2 weeks, clusters of adherent cells morphologically resembling radial glia-like NSCs emerged. After passaging, these rapidly adopted a more uniform morphology and stable proliferation rates. Cell death was minimal and most cultures could be maintained for >10 passages without signs of proliferative arrest or spontaneous differentiation. Expandable neocortex, striatum and brainstem-derived cultures were consistently established from all three specimens, while cultures from other regions did not expand long-term in some cases. RNA was collected from two independent passages of each established line for mRNA profiling (Fig.1). The H3.3-G34R-mutant pGBM002 cell line was derived from a biopsy of a hemispheric tumour in a 15-year-old patient and kindly provided by Angel M. Carcaboso (Institut de Recerca Sant Joan de Deu, Spain)

Cell line engineering

Human foetal NSCs were engineered at 5-7 passages after establishment. Transfections were performed using the Amaxa 4D nucleofector system as previously described (Bressan et al. 2017). For H3.3 expression, 200ng of PiggyBac donor plasmid (WT, K27M- or G34R-H3.3) was co-transfected with 400ng of PBase transposase plasmid. For combined PDGFRA expression and *TP53* knockout, the PiggyBAC PDGFRA donor vector (200ng), *TP53* sgRNAs (200ng each), *TP53* targeting vector (200ng) and Cas9 encoding plasmid (400ng) were added to the reaction. Following recovery, cells were selected for stable integration of PiggyBac donor vectors and *TP53* targeting vector using sequential treatment with 50μg/ml hygromycin (H3.3), 5 μg/ml blasticidin (PDGFRA) and 100 ng/ml puromycin (*TP53* knockout). The resulting resistant cell population was expanded for 3-4 passages before functional and transcriptional profiling. For knockout of the H3.3-G34R allele, a customised allele-specific crRNA was hybridised with a universal tracrRNA (both from Integrated DNA Technologies) and complexed with recombinant Cas9 protein as previously described (Dewari et al. 2018). The ribonucleoprotein complex was delivered into pGBM002 cells using the Amaxa-4D nucleofector system (Lonza, SG solution, program EN137). Following recovery, cells were plated at low density (approx. 10 cells/cm^2) and individual colonies manually picked after three weeks. Clones were expanded and the *H3F3A* locus amplified by PCR and sequenced by Sanger sequencing. *FOXG1* knockout was performed through CRISPR/Cas9-mediated gene targeting using previously validated sgRNA and targeting vectors (Bulstrode et al. 2017). Drug resistant colonies were manually picked and screened by PCR-based genotyping and Sanger sequencing as previously described (Bressan et al. 2017). For xenotransplantation experiments, engineered lines were co-transfected with the Luciferase-2A-GFP donor (200ng) and PBase transposase (400ng), and GFP-positive cells enriched by FACS.

Cell line maintenance

Established lines and engineered derivatives were propagated under identical culture conditions in serum-free basal medium supplemented with N2 and B27, Laminin-1 (1 μg/ml) and growth factors EGF and basic FGF (10 ng/ml each) as previously described (Bressan et al. 2017). Medium was exchanged twice per week. Upon confluence, cells were dissociated with Accutase solution and split typically 1:3 to 1:5.

Immunocytochemistry

Cells were fixed in 4% paraformaldehyde (PFA), permeabilized in 0.1% Triton X-100 and blocked in 0.1% bovine serum albumin plus 3% goat serum PBS solution. Samples were incubated overnight at 4°C with primary antibodies, washed in 0.3% Triton X-100 and reincubated with appropriate fluorophore-conjugated secondary antibodies (1:1000) and 4′,6-diamidino-2-phenylindole (DAPI) at room temperature for one hour. Immunopositive cells were quantified from a total ~5000 cells in a minimum of ten random fields. Total cell number was determined by DAPI nuclear staining. The following primary antibodies were used: active Caspase3 (1:500), Foxg1 (1:10), H3K27me3 (1:1000), H3.3 (1:200), H3.3-G34R (1:1000), Nestin (1:500), Sox2 (1:50), V5-tag (1:1000) and Vimentin (1:50).

Western immunoblotting

Whole cell protein lysates were prepared in RIPA buffer (25 mM Tris-HCl. pH7.6, 150 mM NaCl, 1% NP-40, 1% Sodium Deoxycholate, 0.1% SDS) containing 1x protease inhibitor cocktail. Briefly, the cell pellets were resuspended in 100ul of RIPA buffer, incubated 5 minutes on ice and centrifuged at 4°C (10 minutes 15,000g) to remove cell debris. Histone extracts were prepared as previously described (Shechter et al., 2007). Following protein quantification with the BCA method, extracts were separated on 4–12% Bis-Tris gels and transferred to nitrocellulose membranes using the Mini Trans-Blot Cell system (Biorad). Membranes were incubated overnight with primary antibodies diluted in 5% milk TBS-Tween solution followed by incubation with horseradish peroxidase-conjugated secondary antibodies (1:5000). The following primary antibodies were used: Foxg1 (1:50), GAPDH (1:1000), p53 (1:500), V5 tag (1:1000), total H3 (1:2000), H3.3 (1:1000), PDGFRA (1:1000), p21 (1:1000), ATRX (1:100), Zmynd11 (1:500), HIRA (1:1000), H3K9-K14-K18-K27ac (1:1000), H3K27ac (1:1000), H3K4me2 (1:1000), H3K4me3 (1:1000), H3K9me3 (1:1000), H3K27me2 (1:1000), H3K27me3 (1:1000), H3K36me2 (1:5000) and H3K36me3 (1:1000).

EdU-incorporation assay

Cells were plated at 5,000 cells/cm^2 in 48-well plates for 72h, followed by incubation with 10μM EdU for, unless otherwise stated, 24h. EdU staining was carried out using the Click-iT EdU assay kit according to manufacturer's instructions. EdU-positive cells were quantified from a total

~5000 cells in a minimum of ten random fields. Total cell number was determined by DAPI nuclear staining.

Colony formation and confluence analyses

For colony formation assays, cells were plated at 15 cells/cm2 in 6-well plates, and colony counting performed manually 21 days after plating. For confluence analysis, cells were plated at 5,000 cells/cm^2 in 24-well plates and imaged on the Incucyte Zoom live cell imaging system (Essen Bioscience) until confluence was reached. Growth curves were generated with the built-in Incucyte Zoom software.

SA-βGal staining

Cells were plated at 5,000 cells/cm^2 in 48-well plates for 96h, followed by fixation with solution containing 2% formaldehyde and 0.2% glutaraldehyde for 5 min. Staining was performed overnight with the Senescence Cells Histochemical Staining kit according to manufacturer's instructions. Plates were imaged using phase contrast microscopy and positive cells manually quantified from a total ~2000 cells in a minimum of ten random fields.

Xenotransplantation

Xenotransplantation experiments were performed as previously described (Pollard et al. 2009). Briefly, cells were detached with Accutase, spun down and diluted at 100,000 cells per microliter in sterile PBS. Two microliters of the cell solution were injected stereotactically into the striatum of 6- to 8-week-old NSG mice under general anaesthesia. For luciferase imaging experiments, D-luciferin (50mg/kg dose) was injected intraperitoneally and animals imaged 20 minutes after with the PerkinElmer IVIS Lumina LT Series III (PerkinElmer) instrument. Animals were observed regularly until they became moribund or showed severe neurological symptoms, at which point they were sacrificed and the presence of intracranial tumours confirmed under a fluorescent stereoscope. All procedures were carried out in a designated facility under licenses issued and approved by the UK Government Home Office

Histological analysis

Brain tissue was collected and fixed in 4% PFA solution overnight. Coronal sections (50µm) were processed with the Vibratome instrument (VT1000S, Leica) and blocked in PBS containing 0.2% Triton X-100 and 10% goat serum for 1h at room temperature. Sections were incubated overnight at 4°C with an anti-GFP antibody (1:500) followed by incubation with fluorophore-conjugated secondary antibody (1:1,000; Invitrogen) and DAPI for 1h at room temperature. Slides were mounted using FluorSave reagent and examined with a confocal microscope (TCS SP8, Leica).

RT-qPCR

Total RNA was extracted using the RNeasy Plus Mini Kit following manufacturer's instructions and cDNA generated with random hexamers and SuperScript III Reverse Transcriptase. Quantitative PCR was performed using universal TaqMan PCR master mix and commercial Taqman gene expression assays.

RNA-sequencing

Total RNA was extracted using the RNeasy Plus Mini Kit following manufacturer's instructions. RNA quality was confirmed using the Bioanalyzer RNA 600 Nano kit (Agilent). Libraries were generated with 200ng of RNA input using the KAPA mRNA HyperPrep kit and KAPA SeqCap adapters. Fragmentation was carried out at 94°C for 6 minutes and libraries amplified for 12 cycles. Pooling ratios for multiplex library sequencing was calculated according to individual library concentrations and size distributions assessed with the Qubit dsDNA high-sensitivity and Agilent Bioanalyzer DNA 1000 assays, respectively. Pooled libraries were diluted and processed for sequencing using the Illumina HiSeq 2500 platform.

Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq)

Cells were fixed in culture media containing 1% formaldehyde at room temperature for 10 minutes. The crosslinking reaction was quenched by addition of glycine to a final concentration of 0.125M to the fixation solution for 5 minutes at room temperature. Fixed cells were detached through mechanical scrapping, pelleted by centrifugation and washed twice with ice-cold PBS. The cell pellet was resuspended in LB1 buffer (50 mm HEPES, 140mm NaCl, 1 mm EDTA, 10% glycerol, 0.5% NP40 and 0.25% Triton X100) followed by incubation at 4C for 10 minutes. A nuclear fraction was subsequently obtained through centrifugation and resuspended in LB2 buffer (10 mM

Tris ph8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA). Following centrifugation, the pelleted nuclei were lysed in ChIP sonication buffer (50mM HEPES, 140mM NaCl, 1mM EDTA, 1mM EGTA, 1% Triton X-100, 0.1% Sodium Deoxycholate, 0.3% SDS) and sonicated in a Covaris M220 equipment to yield 200–500 bp DNA fragments. Protease inhibitor cocktail was added to all LB buffers. ChIP-grade antibodies (10ug per reaction) were coupled to Dynabeads Protein G (50ul per reaction) for 6h at 4°C, and added to sonicated chromatin for 16h under gentle rotation at 4°C. Beads were precipitated using a magnetic rack and washed five times with wash buffer (50mM HEPES, 500mM NaCl, 1mM EDTA, 1% NP40 and 0.7% sodium deoxycholate). Immunoprecipitated DNA was eluted in 50mM Tris-HCl pH8.0, 10mM EDTA, 1% SDS for 60 minutes at 65ºC. Following an overnight incubation at 65ºC for crosslinking reversal, samples were treated with RnaseA (0.2 mg/ml) and proteinase K (0.2 mg/ml), and DNA fragments subsequently purified with MinElute PCR purification kit. Purified DNA was quantified using a Qubit fluorimeter, and 20 ng was used to generate ChIP-seq libraries using NEBNext Ultra II DNA Library Kit and NEBNextdual index primer sets for Illumina. Library DNA was quantified using the Qubit, and size distributions were ascertained on a Tapestation (Agilent) using the D1000 ScreenTape assay. Libraries were pooled at equimolar ratio multiplex library sequencing using the concentration and size distribution values of the individual libraries. The library pool was diluted and processed for sequencing on the Illumina NovaSeq S2 50PE platform in accordance with the manufacturer's instructions.

V5-H3.3 Immunoprecipitation

Cells were detached with accutase solution, washed in PBS and frozen at -80C. Typically two confluent 15cm dishes were used per line or replica. Cell pellets were thawed and resuspended in 1ml of lysis buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5% NP-40, 5% glycerol, 1 mM DTT, 2 mM $MgCl_2$ 1mM NaF, 2 mM ortovanadate and protease inhibitor cocktail. Following resuspension, benzonase nuclease (150 units/ml) was added and samples incubated for 30 min at 4C in a rotating shaker. Lysates were centrifuged at 20,000g for 10 min at 4C and supernatants collected for protein quantification using the BCA method. Anti-V5 tag magnetic beads were added to the lysates (50ul of beads to 1mg of protein), and samples incubated in a rotator for 4 hours at 4C. Beads were subsequently precipitated using a magnetic rack, and washed three times with the lysis buffer without NP-40. For western blotting-based experiments,

beads were subsequently resuspended in 50ul of LDS buffer, incubated at 95C for 10min, and separated on 4–12% Bis-Tris gels as described above. Input controls consisted of 5ul of the lysates prior addition of anti-V5-tag beads. Alternatively, washed beads were flash frozen in liquid nitrogen and stored at -80C until further processing for mass-spectrometry as below.

Mass-spectrometry

After washing, beads were incubated for 30 minutes at 27C in 2M urea, 75 mM Tris pH8.5, 1mM DTT, and 0.3 ug trypsin in a total volume of 100uL. The supernatant was incubated 37C overnight and alkylated by addition of 8uL 0.5M iodoacetimate with 20 minutes incubation in the dark. Samples were acidified by addition of 8uL 10% TFA and spun at 20000g. Peptides were loaded onto activated (15 uL methanol) equilibrated (50 uL 0.1 % TFA) stage tips, and washed with 50 uL 0.1 % TFA. Elution in 80% acetonitrile 0.1 % TFA was followed by solvent removal in a vacuum concentrator, and volume then adjusted to 15 uL with 0.1 % TFA. Peptides were loaded onto a C18 packed emitter and separated on a 40 minute gradient from 3.6 % to 79 % acetonitrile with 0.5 % acetic acid throughout. Data dependent acquisition was performed on a Q Exactive Plus with precursor scan 300-1650 70k resolution and fragmentation NCE 28. Data was processed with MaxQuant v1.6.3.3 and Uniprot human proteome release 2018_09. Further technical details and raw data are available at ProteomeXchange with identifier PXD021270.

Bioinformatics Methods

<u>RNA-seq pre-processing and analysis</u>

RNA sequencing reads were trimmed with Cutadapt (version 1.14) with default parameters and aligned to hg38 genome using the pseudo aligner Kallisto (version 0.43.1) (Martin, 2011; Bray et al., 2016). Read counts were summarised via the package Tximport (version 1.8.0) (Soneson et al., 2016) and subsequent normalisation was completed via DESeq2 (version 1.27.32) (Love et al., 2014). Reported normalised read counts, unless otherwise stated, are as regularised-logarithm transformation (rlog) and are used for all downstream principle component analysis (PCA), clustering and visualisation (Love et al., 2014). Differential expression was completed with DESeq2 with $\log_2$ fold change (LFC) shrinkage (Love et al., 2014). GO enrichment analysis was completed with the R package clusterprofiler (version 3.15.4), using the Gene ontology (GO)

database for Biological process (BP), where genes with non-zero read count were used as background (Harris et al., 2004; Yu et al., 2012).

Derivation of forebrain hindbrain specific signatures

Genes specific to forebrain or hindbrain tissue samples were identified via the Distal binarization approach as previously described (Corces et al., 2018). In brief, intra-group mean and standard deviations (SD) were calculated for each gene. Each gene group was then ranked by mean expression across replicates. For any gene where the mean expression of a group is higher than 3 standard deviations than the group preceding it in rank, this gene is labelled specific to this sample group. All samples ranked higher than this group based on gene expression were also labelled as gene specific. Genes specific to forebrain or hindbrain samples were defined as genes specific to only these samples and no other groups. Three SD was chosen as a stringent threshold to accommodate low sample sizes. Pre-filtering of genes was completed using DESeq2 Likelihood Ratio test to filter out genes that were not significantly different across groups (FDR < 0.05). Genes were further filtered such that only specific genes that were also differentially expressed (FDR < 0.05, LFC > 2), between forebrain and hindbrain, were retained. This process was repeated in cell line samples and a concise forebrain and hindbrain set was derived from the intersection of these two procedures.

Derivation of regional signatures in cell line samples

Regional signatures within forebrain (neocortex, striatum, diencephalon) and hindbrain (anterior/posterior brainstem, and spinal cord) NSC cultures were derived by completing differential expression between each individual group against all others using DESeq2 (FDR < 0.05). Genes with low read count (< 1 read) across all samples were pre-filtered. Each set was permitted to overlap, and only the top 50 genes in each comparison, ranked by LFC, were chosen. Distal binarization was not utilised as samples were not distinct enough between each subgroup. Cerebellum and Midbrain signatures were not derived due to low or absent sample numbers (n=1, n=0 respectively). Anterior brainstem and posterior brainstem samples were merged due to high similarity.

Affymetrix microarray pre-processing and analysis

Affymetrix Human Genome U133 Plus 2.0 Array CEL files were downloaded for 3 separate studies from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) (GEO accessions: GSE34824, GSE36245, GSE19578) and re-processed using the oligo package (version 1.51.2), and rma normalisation in R (Carvalho et al., 2010). Pre-filtering of probes was completed by keeping probes with the highest variance across samples per gene annotation. Differential expression was completed using limma (version 3.43.11). Using knowledge of H3.3 mutations, batch correction was completed using the removeBatchEffect function in limma prior to data visualisation, clustering and PCA (Ritchie et al., 2015).

Derivation of pHGG G34 and K27 mutant signatures

G34 and K27 signatures were derived by completing differential expression between each H3.3 mutant group against all others (FDR < 0.1, top 50 genes). A higher FDR threshold was chosen to ensure a comparable gene set size to prior gene sets. As above differential expression was completed on rma normalised intensities using the package limma (Ritchie et al., 2015).

Characterising samples with derived signatures

Principal component analysis (PCA) was completed on pHGG samples using batch corrected rma normalised intensities, and NSC lines using normalised reads (rlog). Reciprocal gene sets were used in each PCA analysis whereby specific forebrain/hindbrain gene sets were used for pHGG samples, and G34/K27 gene sets were used in NSC samples. Reciprocal ssGSEA was completed on each sample set using forebrain/hindbrain, G34/K27, and additional regional signatures (neocortex, striatum, diencephalon, anterior/posterior brainstem and spinal cord) (Barbie et al., 2009). For microarray and RNA-seq data ssGSEA was completed using the R package gsva (version 1.35.7), with the ssgsea method (Hänzelmann et al., 2013), on batch corrected rma intensities, and rlog normalised reads respectively. K-means consensus clustering was completed using the same samples and signatures described above with the R package ConsensusClusterPlus (version 1.52.0) (Monti et al., 2003; Wilkerson et al., 2010). Consensus clustering was run with 10000 re-samples, a proportion of item and feature sampling of 0.7, with maximum k clusters of 10 and 6 for NSC and pHGG samples respectively. Optimal k-clusters were chosen using the CDF (cumulative density function) of the consensus index across varying k clusters, and by inspecting

the consensus matrix for each value of k. Consensus heatmaps were clustered by hierarchical clustering using Euclidean distance with average linkage criteria.

Pediatric cBioPortal gene expression data

Normalised expression (z-score) values of selected genes of interest were retrieved from the paediatric cBioPortal (https://pedcbioportal.kidsfirstdrc.org) using the Paediatric High Grade Glioma (ICR London, Cancer Cell 2017) study. Annotation of tumour location and H3.3 mutation status was retrieved together with gene expression values. Statistical analyses were performed using GraphPad Prism software.

ChIP-seq pre-processing and analysis

ChIP-seq reads were quality and adapter trimmed with Cutadapt, before being aligned to the human hg38 genome with bowtie2 (version 2.3.5) using default parameters (Martin, 2011; Langmead et al., 2012). Resulting BAM files were sorted and indexed with samtools and reads were filtered from ENCODE blacklist regions with samtools view (version 1.9) and bedtools intersect (version 2.29.0) (Li et al., 2009; Quinlan, 2014). PCR duplicates were filtered with picard tools MarkDuplicates (version 2.18.26) (Broad Institute, 2020). Peaks were called on filtered input and control BAM files with MAC2 set to broadPeak (H3.3, H3K36me3, H3K27me3), and narrowPeak (H3K4me3) (Zhang et al., 2008). Narrow peaks were called using MACS2 (version 2.1.2) with the following command line parameters: macs2 callpeak -t Input.bam -c Control.bam -f BAMPE -g 2701495761 -s 50 --keep-dup all -B --SPMR --nomodel --shift 0 --extsize 146 -q 0.05. Broad peaks were called with the additional parameters --broad --broad-cutoff 0.1 instead of -q 0.05. RPM normalised bedgraph files were generated via MACS2 following peak calling. Subtracted coverage files were generated using MACS2 bdgcmp between input and control BAMs. BigWig coverage files were generated using bedGraphToBigWig (version 4) from MACS2 resulting bedgraphs (Kent et al., 2010). Where merged samples are reported, samtools merge was used on filtered BAMs across sample replicates before undergoing the same peak calling, bedgraph, and bigWig generation process.

Heatmap and metagene plots were generated with deeptools (version 3.3.1) plotHeatmap and plotProfile respectively using subtracted bigWig files where negative values were set to 0 (Ramírez

et al., 2014). Where deeptools metagene plots were completed across gene features these were based on the UCSC hg38 knownGene transcript model as defined by the TxDb.Hsapiens.UCSC.hg38.knownGene R package (version 3.4.0) (See Sequence feature analysis) (Team BC, 2016). Downregulated and Upregulated gene sets were defined as those significantly differentially expressed (FDR < 0.01) between pGBM002 parental and G34R-KO pGBM002 samples changing in the positive or negative direction. Expressed genes were defined as those with a mean RPKM < 1 across pGBM002 samples.

Coverage of each histone mark over gene regions or genomic bins, for boxplots and scatterplots, were summarised using bedtools multicov on filtered and processed BAM files, and normalised to RPKM using gene or bin length. Histone mark $\log_2$ fold enrichment was calculated by taking the $\log_2$ fold change of RPKM for a given region between pGBM002 parental and G34R-KO pGBM002 samples. Quantiles of histone coverage and mRNA expression for each gene were calculated using RPKM normalised values.

Peak overlaps between samples, and annotation, were completed using the R package ChIPpeakAnno (version 3.22.0) (Zhu et al., 2010). Promoter specific peaks were defined as those overlapping the transcription start site of a given gene +/- 2kb. Gene body peaks were defined as those peaks overlapping any portion of a gene region. Non overlapping shuffled G34R pGBM002 parental peaks were generated using bedtools shuffle (Quinlan, 2014).

Tracks of histone signal were created using IGV (version 2.8.2) with merged and subtracted bigWig files (Thorvaldsdóttir et al., 2013).

Sequence feature analyses

Characterisation of sequence features across the genome were conducted using the UCSC hg38 knownGene transcript model as defined by the TxDb.Hsapiens.UCSC.hg38.knownGene R package. Each gene region was defined by merging all overlapping transcripts for a given gene. For each gene, a single set of exons and introns were likewise defined by merging overlapping exons and introns separately. Transcripts with multiple discreet sites were filtered.

Using this curated annotation GC% for each gene, exon, and intronic region was calculated using bedtools nuc (Quinlan, 2014). CpG island overlap was calculated as the percentage overlap of the CpG islands track from the UCSC genome browser with a given region. Observed-to-expected (o/e) CpG ratio was calculated for each gene, exon, or intronic region as previously defined (Gardiner-Garden et al., 1987). Exon density was calculated by counting exons per transcript, where overlapping exons were merged, and normalising for length. Gene wise exon and intron length were calculated as the sum of the merged exons or introns per gene region.

Downregulated gene sets were defined as those significantly downregulated (FDR < 0.01) between pGBM002 parental and G34R-KO pGBM002 samples and those downregulated genes that overlapped with the "Forebrain Development" GO term. An empirical p-value for each sequence feature comparison between gene subsets and all genes was calculated using a resampling based approach. Where a null distribution was generated by randomly sub-sampling the genome (10,000 iterations), without replacement, for a set of genes the same size as the gene set of interest.

Supplementary Tables

Table 1 | List of genes and associated sample information used for clustering, principal component and gene set enrichment analyses. Related to Figure 1

Table 1 | List of v5-H3.3 (WT and G34R) bound partners identified by IP-MS in patient cell lines pGBM002. Related to Figure 6

References

Abbas, T., and Dutta, A. (2009). p21 in cancer: intricate networks and multiple activities. Nat. Rev. Cancer *9*, 400–414.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell *9*, 1191–1200.

Barbie, DA. et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 462, 108–112.

Bertacchi, M., Romano, A.L., Loubat, A., Mau-Them, F.T., Willems, M., Faivre, L., van Kien, P.K., Perrin, L., Devillard, D., Sorlin, A., et al. (2020). NR2F1 regulates regional progenitor dynamics in the mouse neocortex and cortical gyrification in BBSOAS patients. EMBO J 39(13): e104163.

Bulstrode, H., Johnstone, E., Marqués-Torrejón, M.Á., Ferguson, K.M., Bressan, R.B., Blin, C., Grant, V., Gogolok, S., Gangoso, E., Gagrica, S., et al. (2017). Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. Genes Dev. *31*, 757–773.

Bjerke, L., Mackay, A., Nandhabalan, M., Burford, A., Jury, A., Popov, S., Bax, D., Carvalho, D., Taylor, K., et al. (2013). Histone H3.3 Mutations Drive Pediatric Glioblastoma through Upregulation of MYCN Cancer Discovery 3(5), 512-519.

Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527.

Bressan, RB. et al. Efficient CRISPR/Cas9-assisted gene targeting enables rapid and precise genetic manipulation of mammalian neural stem cells. Development. 144, 635-648 (2017).

Bushweller, T.H (2019) Targeting transcription factors in cancer - from undruggable to reality. Nat Rev Cancer 19(11):611-624.

Carvalho, BS. et al. (2010). A framework for oligonucleotide microarray pre-processing. Bioinformatics 26, 2363–2367.

Castro, D.S., Skowronska-Krawczyk, D., Armant, O., Donaldson, I.J., Parras, C., Hunt, C., Critchley, J.A., Nguyen, L., Gossler, A., Göttgens, B., et al. (2006). Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. Dev. Cell *11*, 831–844.

Chan, K., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., James, C.D., Jenkins, R., Sarkaria, J., Zhang, Z. (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev. 27(9): 985–990.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., et al. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS Biol. *3*, e283.

Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. Science. 362, eaav1898 (2018).

Dewari, PS. et al. An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein. eLife. e35069 (2018).

Fang, D., Gan, H., Lee, J., Han, J., Wang, Z., Riester, S.M., Jin, L., Chen, J., Zhou, H., Wang, J., et al.(2016). The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science 352(6291):1344-1348

Filbin, M., and Monje, M. (2019). Developmental origins and emerging therapeutic opportunities for childhood cancer. Nat. Med. *25*, 367–376.

Funato, K., Major, T., Lewis, P.W., Allis, C.D., Tabar, V. (2014) Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation. Science 346(6216): 1529-1533.

Gardiner-Garden, M. and Frommer, M. (1987). CpG Islands in vertebrate genomes. J Mol Biol, 196(2), pp. 261–282.

Gibson, P., Tong, Y., Robinson, G., Thompson, M.C., Currle, D.S., Eden, C., Kranenburg, T.A., Hogg, T., Poppleton, H., Martin, J. et al. (2010) Subtypes of medulloblastoma have distinct developmental origins. Nature 468, 1095–1099.

Goldberg, A., Banaszynski, L., Noh, K., Lewis, P., Elsaesser, S., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions Cell 140(5), 678-691.

Guo, R., Zheng, L., Park, J., Lv, R., Chen, H., Jiao, F., Xu, W., Mu, S., Wen, H., Qiu, J., Wang, Z., Yang, P., Wu, F., Hui, J., Fu, X., Shi, X., Shi, Y., Xing, Y., Lan, F., Shi, Y. (2014). BS69/ZMYND11 Reads and Connects Histone H3.3 Lysine 36 Trimethylation-Decorated Chromatin to Regulated Pre-mRNA Processing. Molecular Cell 56(2), 298-310.

Hänzelmann, S. et al. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics 14, 7.

Harris, MA. et al. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32, D258-61.

Hennika, T., Hu, G., Olaciregui, N.G., Barton, K.L., Ehteda, A., Chitranjan, A., Chang, C., Gifford, A.J., Tsoli, M., Ziegler, D.S., et al. (2017). Pre-Clinical Study of Panobinostat in Xenograft and Genetically Engineered Murine Diffuse Intrinsic Pontine Glioma Models. PLoS One *12*, e0169485.

Jessa, S., Blanchet-Cohen, A., Krug, B., Vladoiu, M., Coutelier, M., Faury, D., Poreau, B., De Jay, N., Hébert, S., Monlong, J., et al. (2019). Stalled developmental programs at the root of pediatric brain tumors. Nat. Genet. *51*, 1702–1713.

Johnson, R.A., Wright, K.D., Poppleton, H., Mohankumar, K.M., Finkelstein, D., Pounds, S.B., Rand, V., Leary, S.E.S., White, E., Eden, C. et al. (2010) Cross-species genomics matches driver mutations and cell compartments to model ependymoma. Nature 466, 632–636.

Jones, C., Karajannis, M.A., Jones, D.T.W., Kieran, M.W., Monje, M., Baker, S.J., Becher, O.J., Cho, Y.-J., Gupta, N., Hawkins, C., et al. (2016). Pediatric high-grade glioma: biologically and clinically in need of new thinking. Neuro. Oncol. *19*, 153–161.

Kent, W. J., Zweig, A.S., Barber, G., Hinrichs, A.S., Karolchik, D. (2010). BigWig and BigBed: Enabling browsing of large distributed datasets. Bioinformatics. (17):2204-7.

Khazaei, S., Jay, N.D., Deshmukh, S., Hendrikse, L.D., Jawhar, W., Chen, C.C., Mikael, L.G., Faury, D., Marchione, D.M., Lanoix, J., et al., (2020) H3.3G34W promotes growth and impedes differentiation of osteoblast-like mesenchymal progenitors in giant cell tumour of bone. Cancer Discovery. 10(12):1968-1987.

Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., et al. (2002). Mutation of *ARX* causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. Nat. Genet. *32*, 359–369.

Konno, D., Iwashita, M., Satoh, Y., Momiyama, A., Abe, T., Kiyonari, H., and Matsuzaki, F. (2012). The mammalian DM domain transcription factor Dmrta2 is required for early embryonic development of the cerebral cortex. PLoS One *7*, e46577.

Langmead, B., Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2'. Nat Methods. 9(4):357-9.

Larson, J., Kasper, L., Paugh, B., Jin, H., Wu, G., Kwon, C., Fan, Y., Shaw, T., Silveira, A., Qu, et al. (2019). Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma and Drives Restricted Changes in Bivalent Gene Expression. Cancer Cell 35(1), 140-155.e7.

Lewis, P.W., Muller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. Science. 340(6134):857-861.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools'. Bioinformatics, 25(16):2078-9.

Love, MI. et al. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Lu, C., Jain, S.U., Hoelper, D., Bechet, D., Molden, R.C., Ran, L., Murphy, D., Venneti, S., Hameed, M., Pawel, B.R., et al. (2016) Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science 352(6287): 844–849.

Lujan, E., Chanda, S., Ahlenius, H., Südhof, T., Wernig, M. (2012). Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells Proceedings of the National Academy of Sciences 109(7), 2527-2532.

Mackay, A., Izquierdo, E., Bjerke, L., Nandhabalan, M., Temelso, S., Marshall, L., Zacharoulis, S., Bridges, L.R., Ng, H.-K., Li, X., et al. (2017). Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. Cancer Cell *32*, 520–537.e5.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17(1), 10.

Mi, D., Carr, C.B., Georgala, P.A., Huang, Y.-T., Manuel, M.N., Jeanes, E., Niisato, E., Sansom, S.N., Livesey, F.J., Theil, T., et al. (2013). Pax6 exerts regional control of cortical progenitor proliferation via direct repression of Cdk6 and hypophosphorylation of pRb. Neuron *78*, 269–284.

Mohammad, F., and Helin, K. (2017). Oncohistones: drivers of pediatric cancers. Genes Dev. *31*, 2313–2324.

Monteuuis, G., Wong, J.J.L., Bailey, C.G., Schimitz, U., Rasko, J.E.J. (2019) The changing paradigm of intron retention: regulation, ramifications and recipes. Nucleic Acids Research 47(22):11497–11513.

Monti, S., Tamayo, P., Mesirov, J., Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. Mach. Learn. 52, 91–118 (2003).

Ono, K., Takebayashi, H., Ikeda, K., Furusho, M., Nishizawa, T., Watanabe K., Ikenake K. (2008) Regional- and temporal-dependent changes in the differentiation of Olig2 progenitors in the

forebrain, and the impact on astrocyte development in the dorsal pallium. Developmental Biology 320(2)-456-468

Pathania, M., De Jay, N., Maestro, N., Harutyunyan, A.S., Nitarska, J., Pahlavan, P., Henderson, S., Mikael, L.G., Richard-Londt, A., Zhang, Y., et al. (2017). H3.3K27M Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas. Cancer Cell *32*, 684–700.e9.

Paugh, B.S., Qu, C., Jones, C., Liu, Z., Adamowicz-Brice, M., Zhang, J., Bax, D.A., Coyle, B., Barrow, J., Hargrave, D., et al. (2010). Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. J. Clin. Oncol. *28*, 3061–3068.

Pevny, L., and Placzek, M. (2005). SOX genes and neural progenitor identity. Curr. Opin. Neurobiol. *15*, 7–13.

Pollard, SM. (2009) et al. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. Cell Stem Cell. 4, 568-580.

Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics. 47:11.12.1-34.

Ramírez, F., Dundar, F., Diehl, S., Gruning, B.A., Manke, T. (2014). DeepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42(W1): W187-W191.

Ritchie, ME. et al. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47.

Salloum, R., McConechy, M.K., Mikael, L.G., Fuller, C., Drissi, R., DeWire, M., Nikbakht, H., DeJay, N., Yang, X., Boue, D., et al. (2017) Characterizing temporal genomic heterogeneity in pediatric high-grade gliomas. Acta Neuropathol Commun 5(1):78.

Shechter, D., Dormann, H.L., Allis, D.C., Hake, S.B. (2007). Extraction, purification and analysis of histones. Nat Protocols. 2(6):1445-57.

Schwartzentruber, J., Pfaff, E., Jacob, K., Sturm, D., Tönjes, M., Konermann, C., Lindroth, A., Jäger, N., Rausch, T., Korbel, J.O., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226–231.

Seoane, J., Le, H.-V., Shen, L., Anderson, S.A., and Massagué, J. (2004). Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell *117*, 211–223.

Siegenthaler, J.A., Tremper-Wells., B.A., Miller, M.W (2008) Foxg1 haploinsufficiency reduces the population of cortical intermediate progenitor cells: effect of increased p21 expression. Cereb. Cortex 18, 1865–1875.

Soneson, C., Love, M.I., Robinson, M.D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000 Res. 4, 1521.

Steinfeld, H., Cho, M.T., Retterer, K., Person, R., Schaefer, G.B., Danylchuk, N., Malik, S., Wechsler, S.B., Wheeler, P.G., van Gassen, K.L.I., et al. (2016). Mutations in HIVEP2 are associated with developmental delay, intellectual disability, and dysmorphic features. Neurogenetics *17*, 159–164.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.-A., Jones, D.T.W., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., Bender, S., et al. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell *22*, 425–437.

Sun, Y., Pollard, S., Conti, L., Toselli, M., Biella, G., Parkin, G., Willatt, L., Falk, A., Cattaneo, E., and Smith, A. (2008). Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. Mol. Cell. Neurosci. *38*, 245–258.

Thorvaldsdóttir, H., Robinson, J.T, Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. (2):178-92.

Valencia, A.M., and Kadoch, C. (2019). Chromatin regulatory mechanisms and therapeutic opportunities in cancer. Nat. Cell Biol. *21*, 152–161.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., Ljungman, M. (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. 24(6): 896–905.

Verginelli, F., Perin, A., Dali, R., Fung, K.H., Lo, R., Longatti, P., Guiot, M.-C., Del Maestro, R.F., Rossi, S., di Porzio, U., et al. (2013). Transcription factors FOXG1 and Groucho/TLE promote glioblastoma growth. Nat. Commun. *4*, 1–16.

Vladoiu, M.C., El-Hamamy, I., Donovan, L.K., Farooq, H., Holgado, B.L., Sundaravadanam, Y., Ramaswamy, V., Hendrikse, L.D., Kumar, S., Mack, S.C., et al. (2019). Childhood cerebellar tumours mirror conserved fetal transcriptional programs. Nature *572*, 67–73.

Voon, H.P.J., Udugama, M., Lin, W., Hii, L., Law, R.H.P., Steer, D.L., Das, P.P., Mann, J.R., and Wong, L.H. (2018). Inhibition of a K9/K36 demethylase by an H3.3 point mutation found in paediatric glioblastoma. Nat. Commun. *9*, 3142.

Weinberg, D.N., Allis, C.D., and Lu, C. (2017). Oncogenic Mechanisms of Histone H3 Mutations. Cold Spring Harb. Perspect. Med. *7*, a026443.

Wilkerson, M.D et al. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. Bioinformatics. 26, 1572–1573.

Wu, G., Broniscer, A., McEachron, T.A., Lu, C., Paugh, B.S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M., et al. (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat. Genet. *44*, 251–253.

Xuan, S., Baptista, C.A., Balas, G., Tao, W., Soares, V.C., and Lai, E. (1995). Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres. Neuron *14*, 1141–1152.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., Li, B., Barton, M., Li, W., Li, H., Shi, X. (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. Nature 508(7495), 263-268.

Yip, D.J., Corcoran, C.P., Alvarez-Saavedra, M., DeMaria, A., Rennick, S., Mears, A.J., Rudnicki, M.A., Messier, C., Picketts, D.J. (2012) Snf2l Regulates Foxg1-Dependent Progenitor Cell Expansion in the Developing Brain. Dev Cell 22(4): 871–878.

Yoshida, M., Suda, Y., Matsuo, I., Miyamoto, N., Takeda, N., Suratani, S., Aizawa, S. (1997). Emx1 and Emx2 functions in development of dorsal telencephalon. Development 124(1): 101-111.

Young, F.I., Keruzore, M., Nan, X., Gennet, N., Bellefroid, E.J., and Li, M. (2017). The doublesex-related Dmrta2 safeguards neural progenitor maintenance involving transcriptional regulation of Hes1. Proc. Natl. Acad. Sci. U. S. A. *114*, E5599–E5607.

Yu, G., Wang, L., Han, Y,. He., Q. (2012) ClusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omi. A J. Integr. Biol. 16, 284–287.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute., J., Johnson, D.S., Bernstein, B.E., Nusbaum., C., Myers, R.M., Brown, M., Li., W., Liu, X.S., (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9(9):R137.

Zhang, Y., Shan, C., Wang, J., Bao, K., Tong, L., Jia, S., (2017). Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. Sci Rep. 7: 43906.

Zhu, L. J., Gazin, C., Lawson, N.D, Pages, H., Lin, S.M., Lapointe, D.S., Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics. 11:237

Zweifel, S., Marcy, G., Guidice, Q.L., Li, D., Heinrich, C., Azim, K., Raineteau, O. (2018) HOPX Defines Heterogeneity of Postnatal Subventricular Zone Neural Stem Cells. Stem Cell Reports 11(3): 770-783.

Figure 1

FIG.1

Figure 2

FIG.2

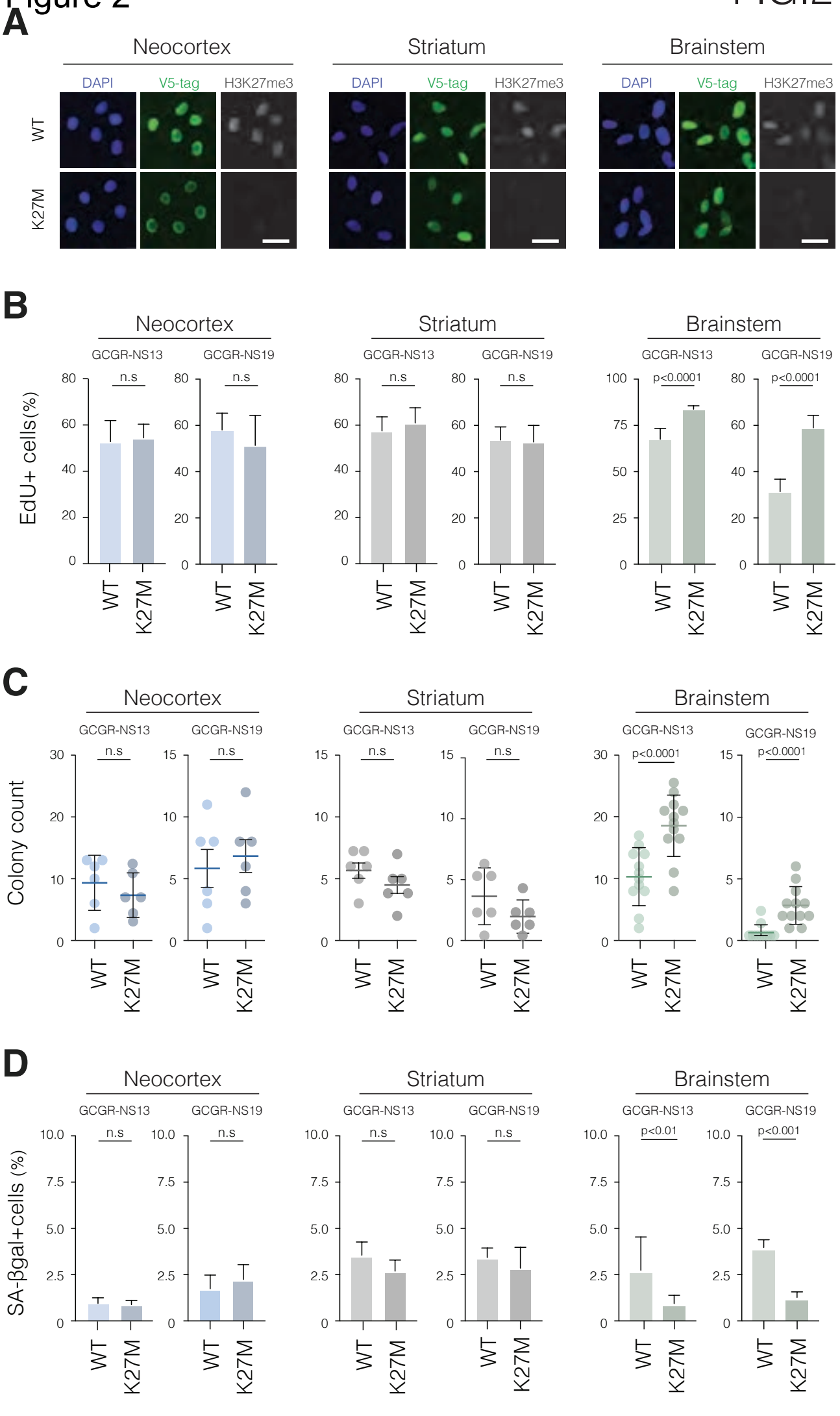


Figure 3

FIG.3

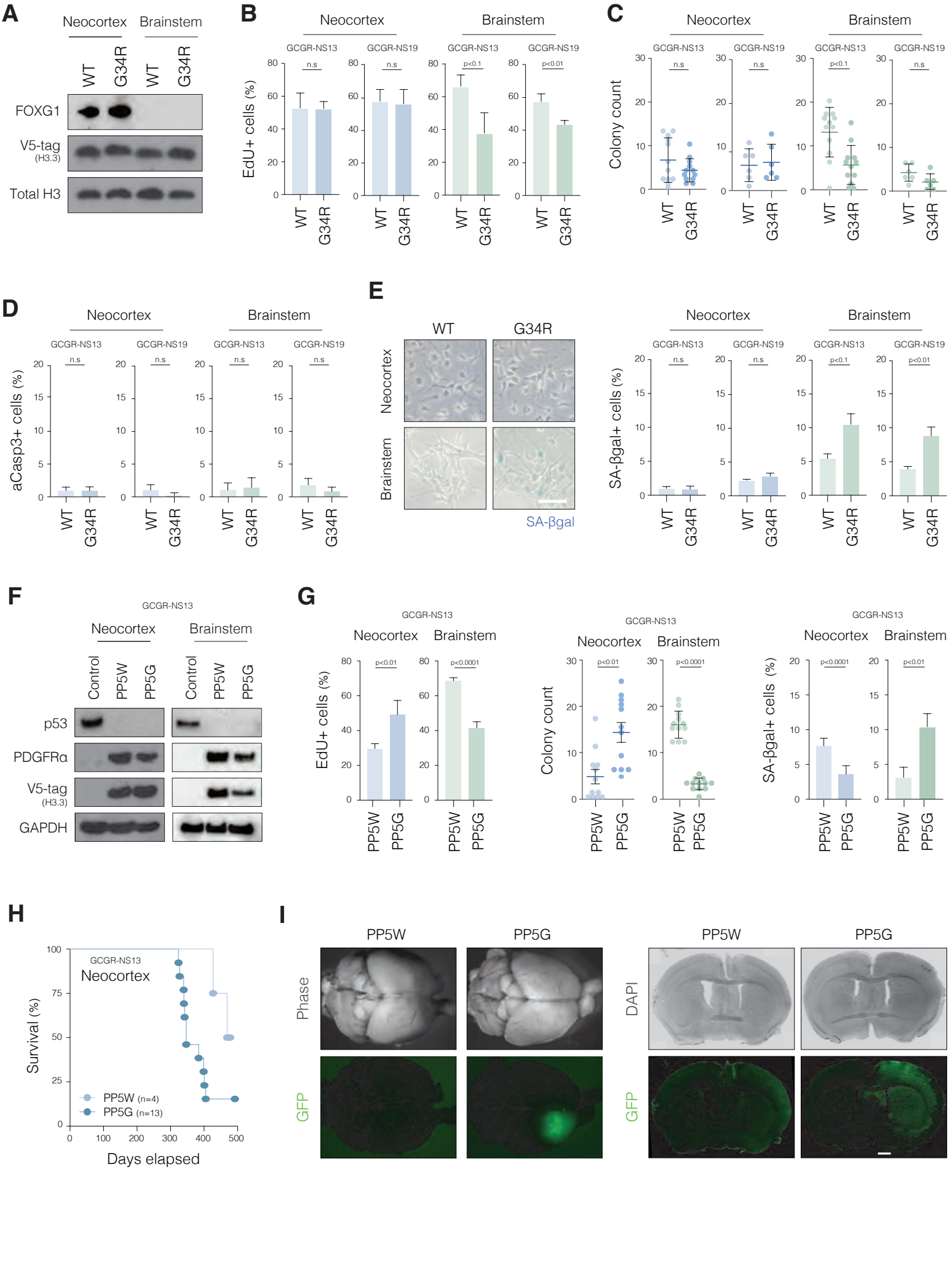


Figure 4

FIG.4

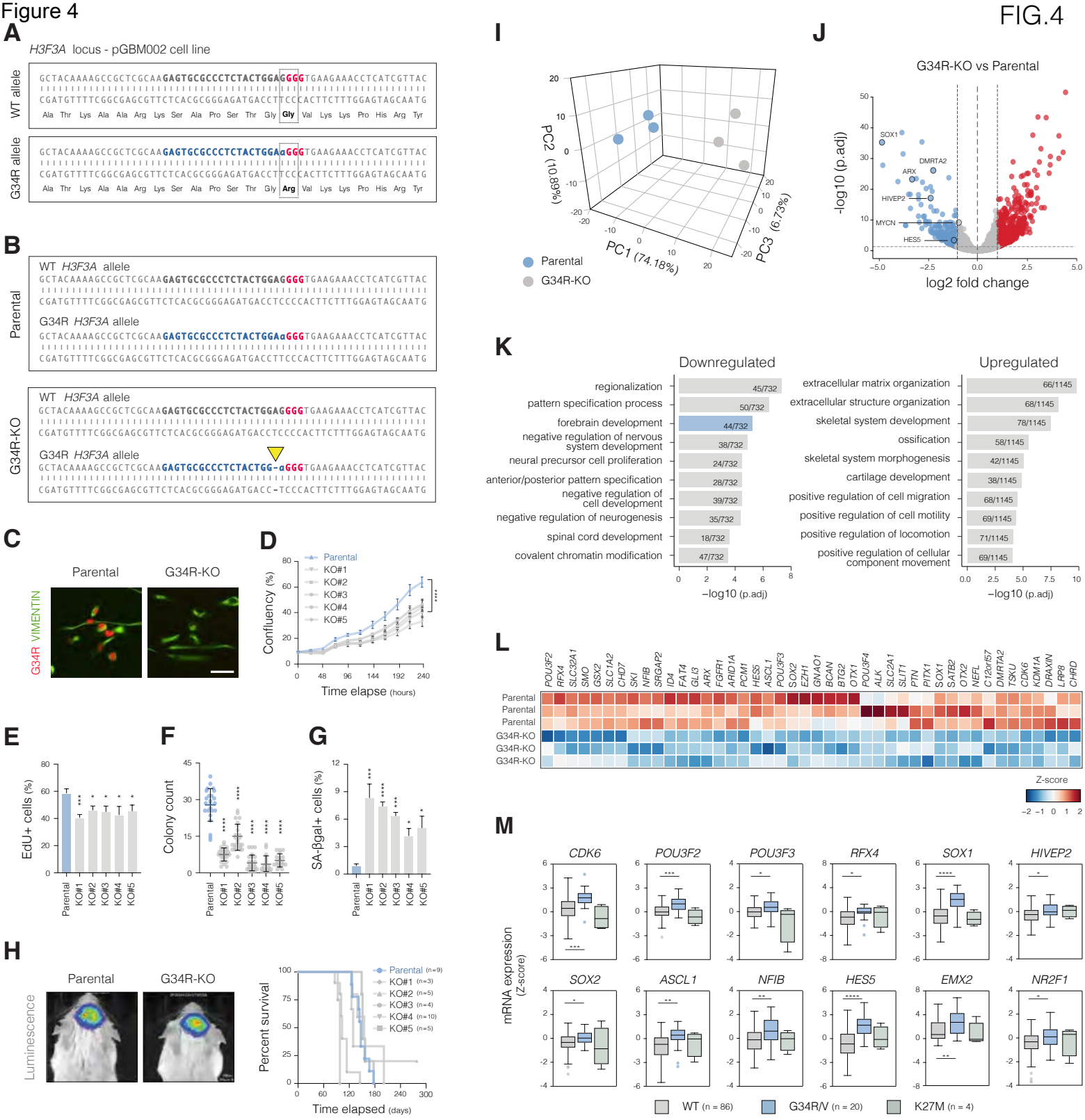
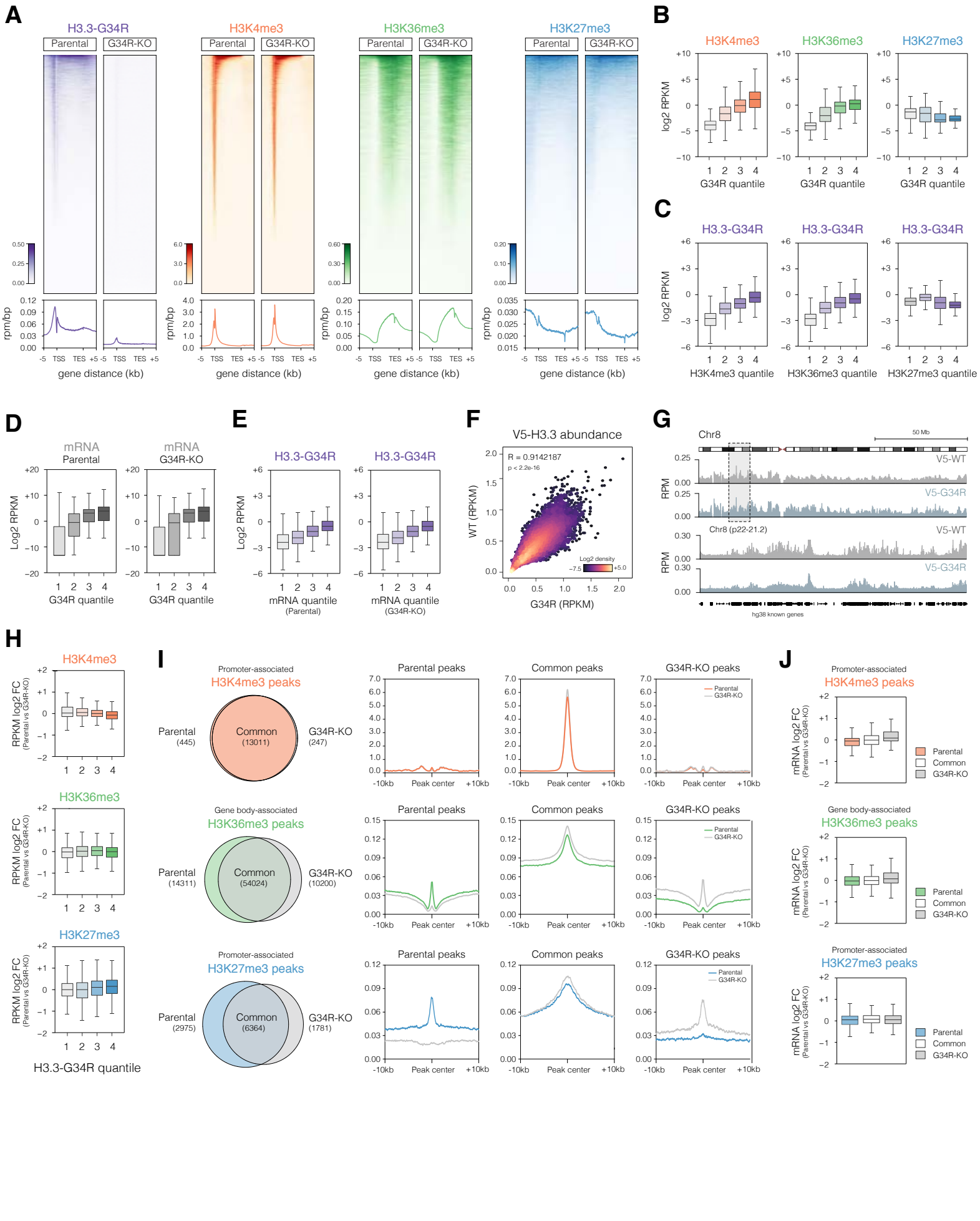


Figure 5

FIG.5



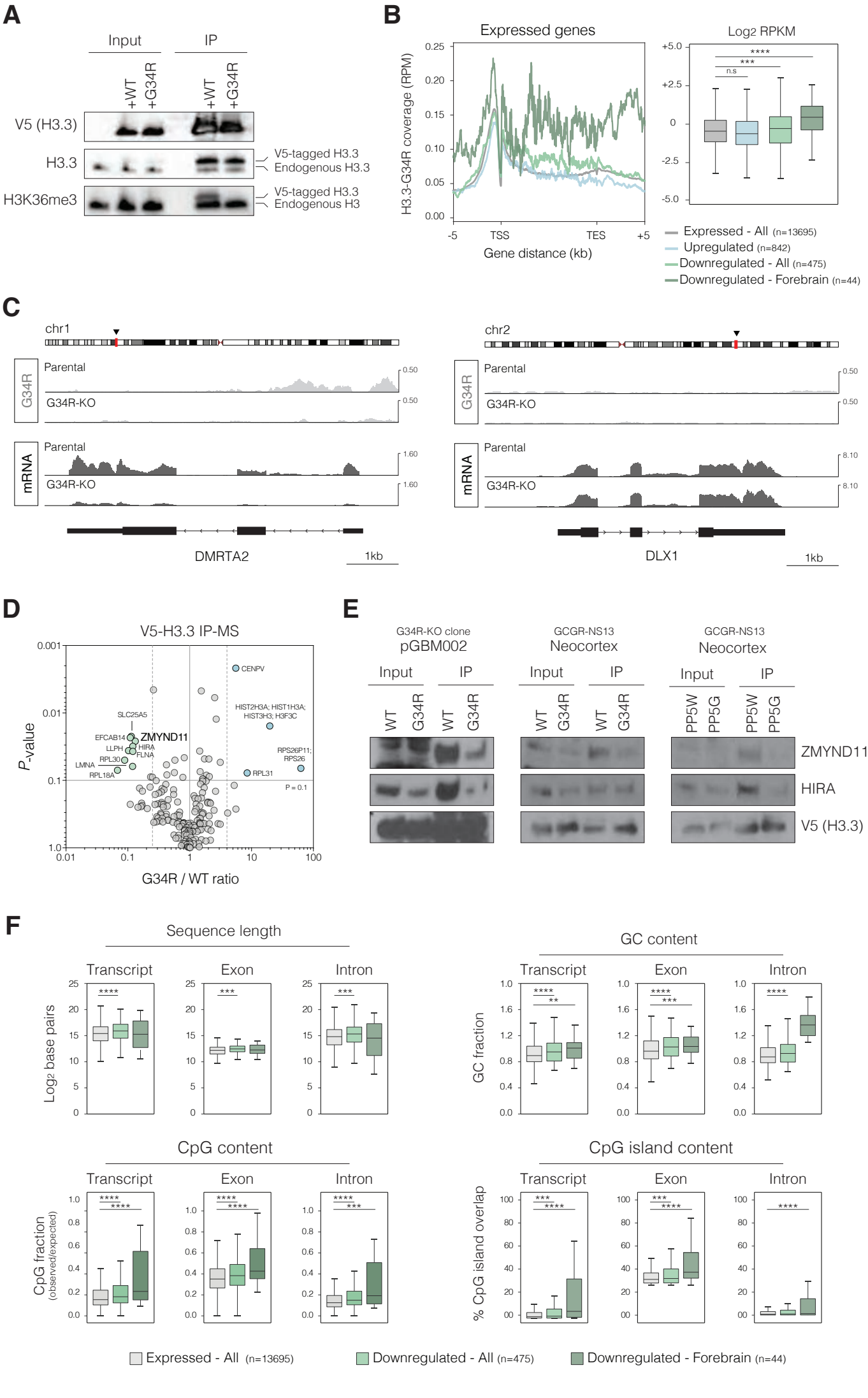

A
Input
IP
+WT
+G34R
+WT
+G34R
V5 (H3.3)
H3.3
V5-tagged H3.3
Endogenous H3.3
H3K36me3
V5-tagged H3.3
Endogenous H3
B
Expressed genes
H3.3-G34R coverage (RPM)
0.25
0.20
0.15
0.10
0.05
0.00
-5
TSS
TES
+5
Gene distance (kb)
Log2 RPKM
+5.0
+2.5
0
-2.5
-5.0
****
***
n.s
Expressed - All (n=13695)
Upregulated (n=842)
Downregulated - All (n=475)
Downregulated - Forebrain (n=44)
C
chr1
chr2
G34R
Parental
G34R-KO
0.50
0.50
mRNA
Parental
G34R-KO
1.60
1.60
8.10
8.10
DMRTA2
1kb
DLX1
1kb
D
V5-H3.3 IP-MS
P-value
0.001
0.01
0.1
1.0
0.01
0.1
1
10
100
G34R / WT ratio
CENPV
HIST2H3A; HIST1H3A; HIST3H3; H3F3C
SLC25A5
EFCAB14
ZMYND11
LLPH
HIRA
FLNA
LMNA
RPL30
RPL18A
RPS26P11; RPS26
RPL31
P = 0.1
E
G34R-KO clone
pGBM002
GCGR-NS13
Neocortex
GCGR-NS13
Neocortex
Input
IP
Input
IP
Input
IP
WT
G34R
WT
G34R
WT
G34R
WT
G34R
PP5W
PP5G
PP5W
PP5G
ZMYND11
HIRA
V5 (H3.3)
F
Sequence length
Transcript
Exon
Intron
Log2 base pairs
GC content
GC fraction
CpG content
CpG fraction
(observed/expected)
CpG island content
% CpG island overlap
Expressed - All (n=13695)
Downregulated - All (n=475)
Downregulated - Forebrain (n=44)

Figure 7

FIG.7